(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) Internati nal Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14589 A2

(51) International Patent Classification⁷: C12Q 1/68

(21) International Application Number: PCT/US00/22769

(22) International Filing Date: 21 August 2000 (21.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,710 20 August 1999 (20.08.1999) US

(71) Applicant: LUMINEX CORPORATION [US/US]; 12212 Technology Boulevard, Austin, TX 78727-6115 (US).

(72) Inventor: CHANDLER, Mark, B.; 4 Niles Road, Austin, TX 78700 (US).

(74) Agents: VILLACORTA, Gilberto, M. et al.; Pepper Hamilton LLP, 600 Fourteenth Street, N.W., Washington, DC 20005-2004 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/14589 A2

(54) Title: LIQUID ARRAY TECHNOLOGY

(57) Abstract: This invention is directed to compositions and methods of screening, sequencing, and/or quantitating a nucleic acid of interest by hybridizing that nucleic acid with a set of oligonucleotide probes, which are coupled to fluorescently addressable multicolored microparticles. These microparticles are provided as a liquid array that can be positioned in predetermined wells or reaction vessels of a microtiter plate. For sequencing purposes, each such liquid array preferably comprises every possible combination of sequences for a given length of a probe. Hybridization occurs by complementary recognition of the analyte of interest with a probe. Probe, target, and/or competing molecule are tagged with a reporter molecule so that upon hybridization, the changes in fluorescence signal parameters are recorded and analyzed.

Liquid Array Technology

CROSS-REFERENCE TO RELATED APPLICATION

The present application claims the benefit of co-pending provisional patent application serial number 60/149,710, filed August 20, 1999, the dislosure of which is incorporated by reference herein. The present invention was funded in part by a Federal Grant from NIST Advanced Technology Program No. 70NANB8H4003. Hence, the Federal Government may have rights to the present invention.

1. FIELD OF THE INVENTION

This invention relates to compositions and methods of manipulating, classifying and obtaining information (including sequence information) on large numbers of nucleic acid molecules. The invention utilizes a plurality of oligonucleotide probes bound to addressable multicolored microparticles that are suspended in a fluid (e.g., a liquid array, suspension array, or gaseous array) and which thus provide at least a three-dimensional approach to carrying out the above-mentioned processes. As such, the present invention represents a radical departure from conventional approaches using "gene chip" technologies, which suffer from the inherent limitations imposed by their two-dimensional confinement.

2. BACKGROUND OF THE INVENTION

There are three basic types of DNA or gene chips (oligonucleotide microarrays). The first and oldest is the sequencing chip. This technology is based on sequencing by hybridization (SBH). A fairly straightforward methodology, it has gained increased utility through the power of computers. For example, an octamer probe constructed of the four basic nucleotides, A, C, T, or G, combined at random (4^8) produces 65,536 possible sequences. By having all possible 65,536 features or probe variants present on a chip and contacting a fragment of deoxyribonucleic acid (DNA) of unknown sequence to the chip, one should obtain useful sequence information about the DNA of interest. It is likely that a DNA fragment of interest will hybridize at various places throughout the chip, and a computer will be able to analyze all areas of overlap and based on that information provide the sequence of the entire DNA fragment. With sequencing chips, such as those initially introduced by Affymetrix and Hyseq, segments of DNA (usually 20 bases or nucleotides long) are placed in a microarray on the surface of a slide. Target samples are then introduced to the chip, and the particular segment(s) that the sample hybridizes with determine(s) the result. For additional discussion on the topic of SBH, see, for example, WO 98/31836 and WO 99/09217, the disclosures of which are incorporated by reference herein. Many other companies are now producing sequencing chips, most using an approach similar to SBH. But whatever their technique, such products are intended to determine the DNA sequence of the sample.

The second variety of DNA chips is known as the expression chip. These are designed to determine the degree of expression of a certain genetic sequence by measuring the rate or amount of messenger ribonucleic acid (mRNA) being produced by the target gene or by measuring complementary DNA (cDNA) corresponding to mRNA. This is done by creating chips with specific sets of base sequences (as opposed to sequencing chips, wherein every possible base sequence is arrayed). Results are then compared to a reference or control, and the degree of change is noted. These chips are useful in diagnosing and treating diseases linked to particular genetic expressions, such as some forms of cancer. Vysis and Synteni are examples of two companies engaged in marketing expression-chip-based products and services. Practical applications and a detailed disclosure of such technology can be found, for example, in Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995 Oct 20;270(5235):467-70 and Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995 Oct 20;270(5235):484-7, both incorporated by reference herein.

The third type of chip is based on comparative or quantitative genomic hybridization. It is designed to help diagnosticians determine the relative amount of a given genetic sequence in a particular sample. A certain amount of unusual genetic expression is normal, but for example in malignant cells the level of expression may be beyond a normal level. Many breast tumors, particularly at the end stages of the disease, are so aberrant genetically that they do not even exhibit the usual number of 46 chromosomes per human cell. (It is known that the 46 human chromosomes house 3 billion base pairs of DNA and encode about 60,000 to 100,000 proteins.) These coding regions on chromosomes make up only about 2% of the genome. The function of the remaining 98% is unknown, and some chromosomes have a higher density of genes than others. This type of chip is designed to look at the level of aberration. This is usually done by using a healthy tissue sample as a reference and comparing it with a sample from the suspected tumor.

Prototype genotyping chips have been originally developed by Affymetrix (Santa Clara, CA, USA), which were used by Wang et al. (Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES, et al., Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998 May 15;280(5366):1077-82) to simultaneously genotype 500 single nucleotide polymorphisms (SNPs). The principles of Affymetrix technology are disclosed in numerous U. S. Pat. Nos. 5,556,752, 5,744,305, 5,837,832, 5,843,655 and 5,874,219 all incorporated herein by way of reference.

Another technology for assaying SNP mutations is being developed by Nanogen (San Diego, CA, USA). Nanogen has developed a semiconductor microchip array that utilizes electrical currents to improve the speed and specificity of binding of nucleotide-specific probes to DNA samples and then assays the bound probes using a fluorescent imaging detector (see, U. S. Pat. No. 5,849,486 to Heller, et al., incorporated by reference herein). This technology allegedly can also perform multiple assays using a variety of different probes simultaneously on a single DNA sample.

Orchid Biocomputer and the University of Washington (Seattle) School of Medicine are establishing a new high-throughput SNP genotyping facility. The center will utilize a proprietary automated genotyping system developed by Orchid, which will enable researchers to sort up to 30,000 SNP genotypes per day for the eventual pharmacogenetic analysis of clinical samples (see, PCT patent publications WO 99/27137 and WO 99/10538 assigned to Orchid, both incorporated by reference herein). Other companies such as Cepheid, Genometrix, Caliper Technologies, Vysis, Genzyme, Gene Logic, Oncor, Myriad Genetics, Jackson Laboratory, Perkin-Elmer and HySeq are developing DNA-array technologies that can potentially contribute to basic research, genome projects, pharmaceuticals, molecular medicine, microbial identification, agriculture, food industry, environment and forensics. A general review on existing chip array technology can be found for example on www.gene-chips.com and the links provided therein; each of which is hereby incorporated herein by reference.

However, the available technology from each of these companies comes as a specifically designed package that includes expensive equipment. Such equipment may include devices for loading samples, imaging, computers, software and modules that will permit different genes to be assayed. It is not possible to mix and match equipment from different manufacturers, so that a given laboratory must use one particular design. If a superior technology is developed, it may be prohibitively expensive to make a change. Therefore, no versatile, simple methods exist to perform chip-based assays.

A microsphere based multiplexed DNA assay, which allows simultaneous analysis of up to 64 unique DNA sequences, is known (see, U. S. Pat. No. 5,736,330 issued to Fulton on April 7, 1998). In this assay oligonucleotide probes complementary to DNA or RNA molecules of interest are coupled to fluorescently labeled microspheres. The probe is allowed to hybridize to analyte nucleic acid in the presence of a competitor molecule, and the amount of the analyte is determined, for example, from the known amount of competitor. By combining just two dyes at eight different fluorescence levels (2^8), one obtains 64 distinct microspheres sets. For example, one unique microsphere set may contain orange and red dyes at a ratio of 85%:15%, respectively, while another unique set may contain a 75%:25% ratio of the same orange and red dyes. In this manner a separate and distinct variable, the level of fluorescence intensity emitted by a given microsphere set, is provided.

3. SUMMARY OF THE INVENTION

According to the present invention, an assay is now provided which allows the resolution of up to 1,000,000 or more unique sets of particles (e.g., microspheres), thus permitting the substantially simultaneous detection of a corresponding number of probes bound thereto. This accomplishment is due in large part to the discovery of methods that allow for the introduction and use of additional dyes, which can be combined at ten or more distinguishable dye ratios giving rise to ten or more unique levels of fluorescence intensity. Hence, for example, the present invention provides for the use of at least six different dyes. The amount of each dye (either singly or in combination with one or more of the other dyes), which is used to stain a specific particle, can be manipulated to provide at least 10 levels of fluorescence intensity. Expressed mathematically, the number of unique sets of particles made possible by such combinations is about 6^{10} or about 1,000,000. The flexibility and versatility of the resulting "fluid array" system allows, in the same system, the pursuit of at least all three possible applications of conventional gene chip technology: That is, sequencing (SBH), detection of expression, and quantification of target nucleic acid.

. What is more the present invention provides superior, novel technology for multiplexed analyses of nucleic acid samples, which technology can be used by virtually any lab using an apparatus no more costly that standard laboratory equipment. Using standard DNA synthetic techniques and equipment based on the principles of conventional flow cytometers, the present invention is able to provide an extremely large array of unique probes. The instant invention is particularly well suited to a multiplexing analysis format and is easily adapted to automated procedures. In short, when using the present fluid arrays, the complexity, expense and limitations associated with prior gene chip manufacturing and/or testing procedures are avoided.

In a preferred embodiment of the present invention a collection is provided comprising multiple subpopulations of particles. The particles in each subpopulation have one or more characteristics that distinguish the particles of one subpopulation from those of another subpopulation, in which the collection is further characterized as having preferably about 10^3 or more distinct subpopulations of particles. In another preferred embodiment, 10^4 or more distinct subpopulations of particles are used. More preferably, the collection is further characterized as having about 10^5 or more distinct subpopulations of particles. An even more preferred collection is one which is further characterized as having about 10^6 or more distinct subpopulations of particles. This collection further comprises bound nucleic acid. Preferably this bound nucleic acid includes a predetermined polynucleotide sequence. Alternatively, this bound nucleic acid includes an ascertainable polynucleotide sequence.

This invention relates to a fluid array which preferably comprises a collection of subpopulations of particles. The particles in each subpopulation preferably have one or more characteristics that distinguish the particles of one subpopulation from those of another subpopulation and are further characterized by a bound nucleic acid. Preferably such a collection is further characterized by having about 10^3 or more distinct subpopulations of particles and a fluid carrier. The preferred fluid carrier is a liquid or gas in which the collection of the present invention is substantially co-mingled with one another, or, if desired, is substantially segregated from one another. The preferred one or more characteristics include a distinctive fluorescence emission signature. The preferred bound nucleic acid that is bound to the particles of one subpopulation differs from that bound to the particles of another subpopulation.

Preferably bound nucleic acid comprises DNA or RNA. The preferred bound nucleic acid comprises 9 or more nucleotide residues. Preferably, 18 or more nucleotide residues are present. Even more preferably 36 or more nucleotide residues are present.

Another preferred composition of the invention comprises a solid particle that has bound to it nucleic acid with a known polynucleotide sequence, a label comprising a dye that exhibits a distinctive fluorescent emission signature, and a substance that, in the absence of an analyte of interest comprising a polynucleotide sequence substantially complementary to said known polynucleotide sequence, can quench the fluorescence emission of the dye.

In the preferred embodiment of the present invention, a device is provided for identifying an analyte of interest, e.g., specific sequence of nucleic acid in a DNA or RNA fragment, among a plurality of unrelated sequences in a sample. This device comprises a fluorescent microparticle having on its surface at least one bound probe of known sequence. Up to one million fluorescently distinct or addressable microparticle sets are contemplated as carrier solid phase substrates for at least the same number of different probes. Non-limiting examples of material for microparticles suitable for use in the present invention include nitrocellulose, nylon, agar, glass, silica, plastic polymer, magnetic beadlets, and the like. Latex or polystyrene microspheres area preferred microparticle. The preferred diameter of microparticles is between about 1 nanometer and about 1,000 microns – a size that distinguishes them from larger size beads employed in the prior art for other applications.

The oligonucleotide probe is constructed to match the desired sequence of the target of interest. Alternatively, a probe is built randomly by selecting any of four bases A, T, C, or G and stacking to these first bases any nucleotide, randomly selected from A, T, C, or G, and carrying the synthesis until the desired probe length is obtained. It is preferable that probes belong to DNA or RNA type molecules (RNA probes will contain as a base uracil or U instead of T), alternatively probes are made of so-called peptide or polyamide nucleic acid (PNA) probes. The probe is synthesized according to standard oligonucleotide synthesis methods using as a solid phase substrate the surface of the microsphere. Alternatively, a probe is synthesized separately and then coupled to the carrier microparticle of choice. In other words the oligonucleotide arrays are fabricated either by *in situ* combinatorial-type synthesis followed by automated sorting of fluorescent carrier particles or by conventional synthesis followed by immobilization of the probe on pre-selected particle sets having unique fluorescence signature.

In this manner a liquid array of up to one million nucleic acid probes is obtained wherein each of the probes is bound to a microparticle stained with at least two fluorescent dyes. Preferably more than one pair of dyes is incorporated into a particle and dyes used in each pair are preferably admixed at variable ratios. Depending on the model of flow cytometer, the preferred number of dyes is between two and six and preferred number of dyes ratio is up to eight or higher.

It is preferable that a probe is labeled with a fluorescent reporter dye, which upon binding of the analyte of interest undergoes through a change in fluorescence attributes indicating the presence of the analyte in a sample and said analyte is identified according to the fluorescent signature of the microparticle. Preferably the operating principle of the fluorescent reporter dye is analogous to one used in TaqMan assay. Other principles of reporting systems are equally suitable and are selected from any of existing methods disclosed in published resources. For example, the target of interest is labeled with a fluorescent tag so that only those that hybridize to a probe are identified. Alternatively, only hybrids are tagged by polymerase extension so that complementary duplexes but not single strand sequences are readily identified. Another way to determine the specificity of hybridization is by providing a competitor molecule that has affinity to the probe.

Accordingly methods for high-throughput screening of nucleic acid samples are provided which are now possible as a result of this liquid array development. The preferred principles of these methods consist of having a labeled probe that matches perfectly or partially with the sequence of interest in a sample. In general, probes used in this invention are not restricted by their length, but preferably they are anywhere between 3 and 120 nucleotides long. More preferably, they are between 6 and 50 bases long. Even more preferably they are between 8 and 25 bases long. In a preferred embodiment, methods of the invention comprise hybridizing to a target nucleic acid a probe having desired length and sequence, wherein the probe hybridizes at various locations, e.g., upstream, within, or downstream of a target nucleic acid.

Another embodiment of the invention is to exploit hybridization process further whereby perfectly hybridized probe will serve as primer for enzymatic extension by polymerase-like enzymes and thus this process is useful for detecting primer-template mismatches, e.g., single nucleotide polymorphism (SNP).

In another embodiment, probes of this invention are used as mixed probes comprising various length oligomers between about 5 and about 120 nucleotides. In this manner the mixture of probes takes advantage of the high selectivity of a short probe and the hybridization stability imparted by a longer probe. Each probe is larger than the first probe but contains the original sequence of the first probe with degenerate additional sequence. Since the genetic code is degenerate (e.g., histidine could be encoded by CAC or CAT), the oligomer probe is prepared with wobbles at the degenerate sites (e.g., for histidine, CAY is used, where Y = C/T). Oligomers with wobbles are also useful in random mutagenesis and combinatorial chemistry. The standard code letters for specifying a wobble are:

R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, and N=A/C/G/T. By requiring hybridization of at least two contiguous probes of different length, false positive signals are reduced or eliminated. As such, the use of varying length oligonucleotides eliminates the need for careful optimization of hybridization conditions for individual probes, as presently required in the art, and permits extensive multiplexing. Several oligonucleotide probes of varying sizes are contemplated as useful to probe several target sequences assayed in the same reaction. The present invention allows up to one million probes to be present simultaneously in the reaction vessel.

An oligonucleotide probe of the invention is hybridized with one or more nucleic acid samples so that complementary complexes are formed. Accordingly, such complexes are analyzed to determine whether the sequence of interest in the complementary complexes match. Complementary sequences are identified by referring to the fluorescent signature associated with each set of microparticles carrying the probe of known sequence. This application is suitable for the SBH-type approach, especially when contiguous probes having predetermined overlapping sequences are used and degree of complementarily or mismatch among overlapping probes reveals the identity of the analyte.

As another embodiment of this invention a liquid array is provided, which comprises a mixture of sets of fluorescently addressable microspheres in a flowable liquid. According to this preferred embodiment each set has a distinct fluorescent signature and each set is conjugated with a different oligonucleotide probe, whereby detection of a fluorescent signature identifies the oligonucleotide probe. This array has no more than 1,000,000 probes of about 9 to 20 nucleotides in length. This array also comprises at least four groups of microspheres, wherein a first group is exactly complementary to a reference sequence and comprises probes that completely span the reference sequence and, relative to the reference sequence, overlap one another in sequence, and wherein three additional groups of microspheres, each of which is identical to said first group but at least one different nucleotide, which different nucleotide is located in the same position in each of the three additional sets but which is a different nucleotide in each set.

In another aspect of this invention detecting the quantity of said analyte is contemplated. Still further, this invention provides means of differential expression of a gene of interest by measuring the levels of mRNA or complementary DNA (or cDNA).

The preferred means of screening are those that are automated including but not limited to flow cytometry, laser scanning microscopy, fluorescence plate reader, and other standard methods known in the art.

A method of detecting at least one mutation in a gene or a set of related genes linked to a clinical condition is provided. Preferably these clinical conditions are associated with genetic aberrations. These aberrations are selected from any mutation types or gene rearrangements, i.e., point mutation, substitution, insertion, deletion, inversion, repetition, amplification, translocation, and transposition. Mutations are either of single nucleotide polymorphism (SNP) type or they involve multiple loci. Large-scale mutations involve chromosome aberrations; smaller mutations involve about 10-50 kbp, and even smaller involve point mutations. Preferably, these mutations are such that they are readily associated with or predisposed to hereditary diseases, neural diseases, muscle and bone diseases, malignant diseases, metabolic diseases, immune diseases, and infectious diseases. Various practical applications are easily envisioned using liquid array assays, which will replace conventional chemical, histological, microbiological, serological and antibody testing methods. Gene based methods of the instant invention are better for: rapid diagnosis of infectious disease; testing latent viral infections; screening of hereditary genetic disease and risk assessment; identification of new therapeutic agents; detection of mutagenic changes to measure genotoxicity of biochemicals; detection of sexually transmitted diseases (STDs) such as gonorrhea, chlamydia, mycobacterium; industry application such as detection of food contaminants, water and sewage testing; identification of genetically transmitted disease; prenatal determination of fetal sex; risk assessment for developing genetically transmitted disease, such as fragile X syndrome, Alzheimer's, Huntington's disease, Gaucher's disease, Marfan syndrome, myotonic dystrophy, diabetes mellitus subtypes, and the like; DNA fingerprinting; forensic testing; paternity/familial testing; cancer diagnosis; agricultural applications; anthropological applications; among many others.

Genomic nucleic acid samples are isolated from a biological sample. Once isolated, the nucleic acids are employed in the present invention without further manipulation. Alternatively, one or more specific regions present in the nucleic acids are amplified. Preferred target amplification techniques include but are not limited to PCR amplification; nucleic acid sequence-based amplification; (NASBA); strand-displacement amplification (SDA); transcription-mediated amplification (TMA); Q-beta replicase; and ligase chain reaction (LCR). Other means of amplification known in the art, e.g., cloning, are equally suitable. An amplification step is not a prerequisite, although it provides the advantage of increasing the concentration of specific nucleic acid sequences within the target nucleic acid sequence population.

The need to contain and manipulate small quantities of reactants is evident. For example, forensic activities often deal with micro-quantities of DNA, bodily fluids, explosives, pesticides, microorganisms, toxins and other residues in trace amounts. Currently, devices exist to facilitate the containment and mixing of small quantities of reactants. Some of these devices include microtiter plates or microwell plates having reaction volumes in the range of 1 to 10 microliters (μl) or higher. However, there are some instances where smaller reaction volumes are required, i.e., between approximately 0.1 and 500 nanoliters (nl) or more. Such small volumes have the capability to contain as little as one microparticle. A preferred microparticle is any micron or submicron particle with size range between about 1 nm up to about 50 μm. In addition, economics often dictate the limited use of reactants. A need exists in the art for a device to contain nanoliter quantities of reactants to facilitate specific interactions between the reactants, and the device must be constructed with readily available materials and also must be easy and economical to use.

In preferred embodiment, the nucleic acid probes are bound to discrete solid-phase supports, i.e., microparticles, suspended as an array in liquid-containing vessels or wells of microtiter plates. In this manner an array of nucleic acid probes is formed wherein each distinct type of a probe is bound to a fluorescently addressable set of microparticles positioned in pre-determined wells of a microtiter plate. This array comprises a plurality of fluorescently addressable microparticles, each stained with at least two fluorescent dyes and is spatially arrayed in a two-dimensional pattern over a plane of a microtiter plate. The identification by location allows the simultaneous processing and screening of a large number of samples. In a preferred embodiment, the microtiter plate has a multiplicity of wells. Depending on the nature of inquiry the number of wells varies, typically 96 to 2,034 wells per plate are suitable. The use of such commercially available plates allows the simultaneous determination of a large number of samples and controls, and thus facilitates the analysis. Moreover, standard automated systems are available to dispense and manipulate reagents in such microtiter plates.

"Plates" are also contemplated which are built specifically for this purpose and are not necessarily similar in their physical appearance to conventional multi-well plates. Considering the submicron scale of fluorescently addressable microparticles, such plates are preferably miniaturized and may fit a surface area that is orders of magnitude smaller than conventional 96-well plate. The preferred minimum contemplated size for a well is one that is sufficient to accommodate a single microparticle, meaning that the total surface area sufficient to accommodate an array consisting of one million microparticles can be smaller than 1mm x 1mm or a surface area smaller than, for example, occupied by letter "o."

Another object of the present invention is to provide a method to screen molecules that bind to nucleic acids representing targets for pharmaceutical intervention. These molecules have various types of biological activities, including, but not limited to, hormonal, neurotransmitter, metabolic, genetic, pharmacologic, immunologic, pathologic, toxic, and anti-mitotic activities. Further, these molecules, e.g., peptides, peptoids, other small organic and inorganic molecules, will be used to design compounds such as drugs, hormones, neurotransmitters, agonists and antagonists more efficiently and economically.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates segmental or combinatorial synthesis and fluorescence sorting of oligonucleotide probes immobilized on uniquely (preferably, fluorescently) addressable multicolored microparticles.

5. DETAILED DESCRIPTION OF THE INVENTION

The inventors have developed an improved process to perform hybridization, enzymatic manipulation, and classification of nucleic acid molecules. The technique utilizes oligonucleotide probes bound to fluorescently addressable microspheres. The identity of the probe is determined by the fluorescent signature of the microparticle and by two-dimensional position of the well in a microtiter plate in which discrete sets of such microparticles are suspended in a hybridization liquid (liquid arrays). The present invention provides arrays of nucleic acid probes immobilized on fluorescent microparticles, methods of making such arrays, and high-throughput screening methods for detecting, measuring, and identifying genetic characteristics of a biological sample. The combination of the above-identified innovations obviates the need for so-called gene or DNA chips. As such, this new technique is a radical departure from conventional gene chip approaches.

The term "complementary" refers to two nucleic acid strands that exhibit substantial normal base pairing characteristics. Complementary strands may contain one or more mismatches, however.

The term "hybridization" refers to the hydrogen bonding-mediated interaction that occurs between two complementary oligonucleotide strands.

The term "multiplex analysis" refers to the simultaneous assay of pooled DNA or RNA samples according to the disclosed methods.

The term "mismatch" means that a nucleotide in one strand of DNA or RNA does not or cannot pair through Watson-Crick base pairing and stacking interactions with a nucleotide in an opposing complementary DNA or RNA strand. Thus, adenine in one strand of DNA or RNA would form a mismatch with adenine in an opposing complementary DNA or RNA strand. Mismatches also occur where a first nucleotide cannot pair with a second nucleotide in an opposing complementary DNA or RNA strands because the second nucleotide is absent (i.e., deleted).

5.1 Flow Cytometry and Library Creation

Flow cytometry (also known as Fluorescence Activated Cell Sorting-FACS) is a powerful method that is used mainly to resolve complex mixtures of cell populations based on their fluorescence characteristics. The target cells are tagged with fluorophores either directly or indirectly by coupling via fluorescently labeled antibodies or ligands. The cells are then pumped in an aqueous stream where vibrational forces create a stream of droplets containing single cells. As the cell-containing droplets pass through the flow cell, lasers excite the fluorophores or fluorescent dyes and the resulting emission is measured in real-time. Up to six colors (six independent fluorescent measurements) are now possible and standard flow cytometers are capable of enumerating cells at flow rates of 10,000-100,000 cells per second in real-time. Advanced flow cytometers also contain electrically coupled deflectors that deflect or gate cells with undesirable fluorescence parameters, thereby permitting the operator to physically separate cells within the population based on fluorescence spectrum and/or intensity.

Taking the advantage of multiplexed flow cytometry based approach described elsewhere (see, U. S. Pat. No. 5,736,330 as incorporated herein in its entirety), a method is developed that combines advantages of FACS and standard nucleotide synthesis chemistry (methods of synthesizing oligonucleotides are found in, for example, Oligonucleotide Synthesis: A Practical Approach, M.J. Gait, ed., IRL Press, Oxford (1984), incorporated herein by reference in its entirety for all purposes). The general scheme and strategy are illustrated in Fig. 1 and are similar in principle to the split synthesis approach in combinatorial chemistry. This strategy is diametrically opposite to the parallel synthesis approach that allows the creation of gene chips.

Sets of fluorescently addressable microspheres are first mixed together. A computer file assigning a unique oligonucleotide sequence to each encoded microsphere is created to control flow sorting. The starting mixture of microparticles is then sorted to 4 outputs of the flow cytometer depending upon which base (A, C, G, or T) is attached first to the surface of the microparticle (various methods of attaching nucleic acid to solid base support are known, e.g., carbodiimide coupling, and details for alternative methods can be found for example in U. S. Pat. Nos. 5,919,626, 5,610,287, and 5,837,860, which are incorporated herein by way of reference). All microspheres that require deoxyadenosine (A) are sorted to output 1 and all those requiring C are sorted to output 2, while G and T are sorted to outputs 3 and 4 respectively. The microspheres are then transferred to standard nucleic acid synthesis machine or automated DNA synthesizer, such as an ABI 392, and appropriate coupling of A, C, G, or T is performed on each of microsphere mixtures. The methods of synthesis are standard methods well known in the art, e.g., U. S. Pat. No. 5,869,644; Southern EM, Case-Green SC, Elder JK, Johnson M, Mir KU, Wang L, Williams JC. Arrays of complementary oligonucleotides for analyzing the hybridization behavior of nucleic acids. Nucleic Acids Res 1994 Apr 25;22(8):1368-73; and Maskos U, Southern EM. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ. Nucleic Acids Res 1992 Apr 11;20(7):1679-84), all of which are incorporated by reference herein. The particles are then pooled and returned to FACS machine and sorting and synthesis are performed repeatedly until desired length of the probe is obtained. Alternatively, another means of obtaining an array is to synthesize in advance the oligonucleotides on an automated DNA synthesizer and then attach the oligonucleotides onto the solid phase (Lamture JB, Beattie KL, Burke BE, Eggers MD, Ehrlich DJ, Fowler R, Hollis MA, Kosicki BB, Reich RK, Smith SR, et al. Direct detection of nucleic acid hybridization on the surface of a charge coupled device. Nucleic Acids Res 1994 Jun 11;22(11):2121-5 and Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Res 1994 Dec 11;22(24):5456-65, both herein are incorporated by reference). It is readily seen that the diversity of obtained sequence or maximum possible number of non-redundant sequences obeys a simple mathematical formula 4^N where N is the number of nucleotides in the oligomer probe and 4 is the number of bases used for synthesis. For example if a probe is a tetramer oligonucleotide then $4^4=16$. Similarly, $4^5=1,024$; $4^6=4,096$; $4^7=16,384$; $4^8=65,536$; $4^9=262,144$; $4^{10}=1,048,576$; $4^{15}=1.073^9$; $4^{20}=1.099^{12}$ and so on.

The method allows massive parallel synthesis to efficiently produce high diversity of oligomers, i.e., oligonucleotide library. Thus, for example, to form an octamer probe array of 4^8 oligonucleotides, only four parallel nucleotide reactions (they are run simultaneously in the same DNA synthesizer) are required at each addition or stacking step (one for each of A, C, G, and T), so that the total for an array or a library consisting of 65,536 discrete sets of probes associated with an equal number of sets of microparticles are produced in just 8 reactions. By contrast, if each probe is made individually, a total of 4^8 separate addition steps would be required. Thus this invention overcomes the complexity of the prior art methods for constructing an oligonucleotide library. Libraries resulting from above-described synthesis approach are characterized by the phrase "one microsphere, one probe." Each microparticle in the library holds at least one and preferably multiple copies of a single library member. Particle-based approach greatly simplifies the isolation and identification of analyte molecules because beads are large enough to be observed by automated means and sorted mechanically.

In addition to construction of probes of random sequence one skilled in the art knows how to create probes of specific affinity to a particular gene or genomic region of interest. Genomic regions suspected to contain one or more mutations are identified by reference to a nucleotide database, such as GENBANK, EMBL, or any other appropriate database or publication disclosing such mutation. If the genomic code is unknown and unavailable in these databases then it is deduced by standard sequencing methods such as disclosed in U. S. Pat. No. 4,962,020, which incorporated herein by way of reference. GENBANK is a computerized database of nucleotide and amino acid sequences that are constantly revised and updated on a daily basis. Although the GENBANK itself originated as genetic sequence database from the National Institutes of Health (NIH), it is not a single entity. Indeed, it is an annotated collection of all publicly available DNA sequences. GENBANK comprises, for example, the International Nucleotide Sequence Database Collaboration, which in turn comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. Protein sequences in the GENBANK are from Swiss-Prot, PIR, PRF, PDB, as well as translated protein sequences from the DNA sequence databases. Additional information is provided by National Library of Medicine through MEDLINE and by the American Type Culture Collection (www.atcc.org). In the preferred embodiment of this invention the content of these databases and updates thereof are incorporated by reference in their entirety. When required, this data is downloaded from Internet to a computer attached to a flow cytometer and desired probe sequences are constructed by using specially designed program according to the combinatorial synthesis strategy described above. This is a significant simplification over the prior art of making "gene chips" which require expensive procedures of photolithographing or masking and ink-jet deposition of nucleic acid probes.

5.2. Sequencing by Hybridization (SBH)

The principle of hybridization and various methods are extensively used in molecular biology and medicine. Methods of performing such hybridization reactions are disclosed by, for example, Sambrook, J. et al. (In: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)), Haymes, B. D., et al. (In: Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985)) and Keller, G. H. and Manak, M. M. (In: DNA Probes, Second Edition, Stockton Press, New York, N.Y. (1993). The general methodology for sequencing by hybridization (SBH) has been described in detail (e.g., U.S. Pat. Nos. 5,770,367 to Southern and Cummins; 5,695,940, 5,667,972, 5,525,464 5,492,806 5,202,231 to Drmanac et al.; 5,902,723 to Dower and Fodor on May 11, 1999; and 5,552,270 to Khrapko et al., which references are incorporated herein by reference). Briefly, the method involves first identifying oligos in an oligo permutation library that bind to the analyte. The sequences of the binding oligos are then reconstructed, by sequence overlap construction, to determine a full analyte sequence. A complete set of 4^N nucleotides of length N is immobilized as an ordered array on a solid support and an unknown DNA sequence is hybridized to this array. The resulting hybridization pattern provides all n-tuple "words" in the sequence. To date, this method has been applied successfully to determination of short sequences, but tends to lead to sequence ambiguities for longer sequences, e.g., 100 Kb or larger. One reason for the sequence ambiguity is duplicated short sequences in the analyte, which can lead to unresolved sequence determination. Although in theory this problem can be solved by generating longer-sequence oligo libraries, this solution requires a much larger and more complex permutation libraries. It also increases the likelihood of single-mismatch hybridization, since longer stretches of duplex will accommodate more base pair mismatching, even under high stringency conditions, especially in duplex regions with high GC content.

Although the invention has been described with respect to particular methods and library arrays, it will be appreciated that various changes and modification can be made without departing from the invention.

A first step is to create a plurality of sets of microspheres with probes or fragments of DNA, wherein fragments are known in advance. The sequence of one fragment or probe may overlap with the fragment of another set of probes. The overlapping of hybridized probe and analyte in a contiguous manner enables one to reconstitute the DNA sequence of interest. Therefore, invented method also provides an efficient means of sequencing of DNA by providing contiguous overlapping probes.

The process of sequentially manipulating molecules involves the use of miniaturized reaction vessels arranged as microtiter plates. Fragments having an average length of about 250 bases are generated by cutting target DNA, with specific restriction endonucleases. These fragments are directly sequenced by the use of contiguous stacking hybridization on a sequencing array.

To effectively separate the fragments from each other, element of an array must contain oligonucleotide strings that are unique for specific fragments. The longer the oligonucleotide string and the shorter the fragmented DNA, the higher the probability that a sequence complementary to the oligonucleotide string will be unique for only one of the fragments. Concurrently, the probability that the oligonucleotide string will hybridize at all with any fragment present in the mixture, will be lowered. Conversely, the shorter the length of immobilized oligonucleotide strings, the higher the hybridization sensitivity to single-base-pair mismatches; however, the stability of the formed duplexes decrease. Furthermore, single-stranded nucleic acids form relatively stable hairpins and tertiary structures that interfere with their hybridization with shorter oligonucleotide immobilized fractions. The introduction of base analogs or the substitution of negatively charged phosphodiester groups in the immobilized oligonucleotides by neutral or even positively charged groups significantly increases duplex stability viz. hairpin stability. For example, substitution of negatively charged phosphate groups for positively charged guanidinium linkages renders the duplex of thymidil 5-mers with poly(rA) stable even in boiling water.

The invented method is also appropriate for drug screening or to construct a protein assay. In one scenario, an array of monoclonal antibodies, heavy and light chains from a spleen library or from polyvalent sera is a suitable source. Then each of these antibodies is immobilized in separate elements of an array. The array is then subjected to an antigen, which is tagged. Those elements that "light up" would serve as starting points for building antibodies specific for that antigen. Further, depending on the size of the array, such arrays of microspheres will serve as a universal antibody diagnostic device allowing for thousands of assays to occur simultaneously via protein affinity processes. Drug screening approach is based, for example, on affinity binding of tested molecules to an array of nucleic acid targets representing a sequence of a gene responsible for a given disease. Chemical and/or pharmaceutical libraries containing potential drugs are tagged with a reporter molecule and are then screened against the array of this invention. Those that bind are identified by reporter molecule or detectable indicator. These compounds are further refined to find optimally the best candidate. Non-limiting examples of DNA-binding molecules are for example estrogen receptor, androgen receptor, thyroid hormone receptor, glucocorticoid receptor, vitamin D receptor, human vascular endothelial growth factor (VEGF), human chorionic gonadotropin (ACG) and human thyroid stimulating hormone (hTSH). Other DNA binding molecules are also contemplated including but not limited to NF-kB binding unit, an SP1 binding unit, a TATA binding unit, a human papillomavirus (HPV) E2 binding unit, an HPV LTR binding unit, and human immunodeficiency virus (HIV) LTR binding unit. Non-limiting examples providing specific technical details and appropriate sequences are found for example in U.S. Pat. Nos. 5,888,738, 5,871,902, and 5,874,218, as incorporated herein by reference.

Base identification for the purpose of sequencing is accomplished according to standard computer "software" programs based on principles such as Bayesian classification algorithm providing variable kernel density estimation, fuzzy logic and neural network, decision tree, rough sets, nearest neighbor techniques, among others. Commercial programs are available but one skilled in the art would know which one is most appropriate in a particular situation and may develop an in-house program. Basically, the likelihood of each identification associated with a set of hybridization values is computed by comparing an unknown set of probes to a set of example cases for which the correct base identification was known.

Methods of the invention permit the detection of a mutation at a locus in which there is more than one nucleotide to be interrogated. Moreover, methods of the invention allow one to screen a locus in which more than one single base mutation is possible. Once regions of interest are identified, at least one probe is prepared to detect the presence of a suspected mutation. This probe is then arrayed with other probes having affinity to other identified mutations so that an array specific to a disease or set of diseases becomes available. Appropriate control probes are included as well, e.g., wild type sequence corresponding to non-mutated normal sequence. The ability to detect mutations in coding and non-coding DNA (exon/intron), as well as mRNA, is important for the diagnosis of inherited diseases. A gene mutation can be a single nucleotide change (SNP) or multiple nucleotide changes in a DNA sequence encoding an essential protein. A single nucleotide change or multiple nucleotide changes can result in frame shift mutations, stop codons, or non-conservative amino acid substitutions in a gene, each of which can independently render the encoded protein inactive. However, a gene mutation can be harmless, resulting in a protein product with no detectable change in function (i.e., a harmless gene polymorphism). Mutations in repetitive DNA can also lead to diseases as is the case, for example, in human fragile-X syndrome, spinal and bulbar muscular dystrophy, and myotonic dystrophy. A mutant nucleic acid that includes a single nucleotide change or multiple nucleotide changes will form one or more base pair mismatches after denaturation and subsequent annealing with the corresponding wild type and complementary nucleic acid. For example, G:A, C:T, C:C, G:G, A:A, T:T, C:A, and G:T represent the eight possible single base pair mismatches which can be found in a nucleic acid heteroduplex, where U is substituted for T when the nucleic acid strand is RNA. Nucleic acid mismatches can form when the two complementary strands of a heteroduplex are derived from DNA or RNA molecules that differ in sequence such that one contains deletions, substitutions, insertions, transpositions, or inversions of sequences compared to the other.

Detection of such mutations provides an important diagnostic tool in areas including cancer diagnosis and prognosis, prenatal screening for inherited diseases, differential diagnosis of diseases not readily detectable by conventional tests (for example, Marfan's syndrome and the fragile X syndrome), and the analysis of genetic polymorphisms or DNA fingerprinting (for example, for genetic mapping or identification purposes in legal and forensic matters).

Methods disclosed herein allow one skilled in the art to detect mutations such as insertions, deletions, and substitutions. Nucleic acid samples to be screened with the methods of the present invention comprise human, animal, plant, and microbial nucleic acid samples. Methods disclosed herein are useful to detect mutations associated with diseases such as cancer. Additionally, methods of the invention are useful to detect a deletion or a base substitution mutation causative of a metabolic error, such as complete or partial loss of enzyme activity. In another embodiment, the specific nucleic acid sequence comprises a portion of a particular gene or genetic locus in the patient's genomic nucleic acid known to be involved in a pathological condition or syndrome. Non-limiting examples include cystic fibrosis, Tay-Sachs disease, sickle-cell anemia, thalassemia, and Gaucher's disease.

Single base mutations can be detected by differential hybridization techniques using allele-specific oligonucleotide probes as disclosed in incorporated reference by Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc Natl Acad Sci U S A 1989 Aug;86(16):6230-4). Mutations are identified on the basis of the higher thermal stability of the perfectly-matched probes as compared to mismatched probes. Nucleic acid hybridization can be generally combined with some other selection or enrichment procedures for analytical and diagnostic purposes.

In enzyme-mediated ligation methods, a mutation is interrogated by two oligonucleotides capable of annealing immediately adjacent to each other on a target DNA or RNA molecule, one of the oligonucleotides having its 3' end complementary to the point mutation. Adjacent oligonucleotide sequences are only covalently attached when both oligonucleotides are correctly base-paired. Thus, the presence of a point mutation is indicated by the ligation of the two adjacent oligonucleotides as disclosed in incorporated reference by Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S, Iovannisci DM, Woo S, Winn-Deen ES. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res 1994 Oct 25;22(21):4527-34). However, the usefulness of this method for detection is compromised by high backgrounds which arise from tolerance of certain nucleotide mismatches or from non-template directed ligation reactions (Barringer KJ, Orgel L, Wahl G, Gingeras TR. Blunt-end and single-strand ligations by Escherichia coli ligase: influence on an in vitro amplification scheme. Gene 1990 Apr 30;89(1):117-22).

A number of detection methods have been developed which are based on a template-dependent, primer extension reaction. These methods fall essentially into two categories: (1) methods using primers which span the region to be interrogated for the mutation, and (2) methods using primers which hybridizes proximally and upstream of the region to be interrogated for the mutation. As used hereinafter the notion primer connotes essentially same meaning as a probe except that after initial binding step as in probe-target binding, the primer serves to direct the synthesis of new strands substantially identical to the template sequence in the amplified product.

In the first category, Caskey and Gibbs in U. S. Pat. No. 5, 578, 458 disclose a method wherein single base mutations in target nucleic acids are detected by competitive oligonucleotide priming under hybridization conditions that favor the binding of the perfectly-matched primer as compared to one with a mismatch. Vary and Diamond in U. S. Pat. No. 4, 851, 331 disclosed a similar method wherein the 3' terminal nucleotide of the primer corresponds to the variant nucleotide of interest. Since mismatching of the primer and the template at the 3' terminal nucleotide of the primer inhibits elongation, significant differences in the amount of incorporation of a tracer nucleotide result under normal primer extension conditions.

It has long been known that primer-dependent DNA polymerases have, in general, a low replication error rate. This feature is essential for the prevention of genetic mistakes, which would have detrimental effects on progeny. Methods in a second category exploit the high fidelity inherent in this enzymological reaction. Detection of mutations is based on primer extension and incorporation of detectable, chain-terminating nucleoside triphosphates. The high fidelity of DNA polymerases ensures specific incorporation of the correct base labeled with a reporter molecule. Such single nucleotide primer-guided extension assays have been used to detect aspartylglucosaminuria, hemophilia B, and cystic fibrosis; and for quantifying point mutations associated with Leber Hereditary Optic Neuropathy (LHON). See. e.g., Kuppuswamy MN, Hoffmann JW, Kasper CK, Spitzer SG, Groce SL, Bajaj SP. Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. Proc Natl Acad Sci U S A 1991 Feb 15;88(4):1143-7; Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 1990 Dec;8(4):684-92; Juvonen V, Huoponen K, Syvanen AC, Nikoskelainen E, Savontaus ML. Quantification of point mutations associated with Leber hereditary optic neuroretinopathy by solid-phase mini-sequencing. Hum Genet 1994 Jan;93(1):16-20; Ikonen E, Manninen T, Peltonen L, Syvanen AC. Quantitative determination of rare mRNA species by PCR and solid-phase mini-sequencing. PCR Methods Appl 1992 May;1(4):234-40; Nikiforov TT, Rendle RB, Goelet P, Rogers YH, Kotewicz ML, Anderson S, Trainor GL, Knapp MR. Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. Nucleic Acids Res 1994 Oct 11;22(20):4167-75, which incorporated herein by way of reference.

The selectivity and stability of the oligonucleotide primer extension assay is determined by the length of the oligonucleotide primer. Under typical reaction conditions, short primers (i.e., less than about a 15-mer) exhibit transient, unstable hybridization and, consequently, do not readily prime the extension reaction. Moreover, in a complex heterogeneous biological sample, short primers may exhibit non-specific binding to a wide variety of perfectly-matched complementary sequences. Thus, because of their low stability and high non-specific binding, short primers are not very useful for reliable identification of a mutation at a known location. Therefore, detection methods based on primer extension assays use oligonucleotide primers ranging in length from 15-mer to 25-mer. See e.g., PCT

Patent Publications WO 91/13075; WO 92/15712; and WO 96/30545. Lengthening the probe to increase stability, however, has the effect of diminishing selectivity. Due to the small thermodynamic differences in hybrid stability generated by single nucleotide changes, a single base mismatch usually does not affect binding efficiency of longer oligonucleotide primers. This tolerance of nucleotide mismatches in the hybridization of the primer to the template can result in significant levels of non-specific false priming in complex heterogeneous biological samples. Methods in the art reduce the possibility of false priming by decreasing the sequence complexity of the test sample. Thus, genomic DNA is isolated from the biological sample and/or amplified with PCR using primers, which flank the region to be interrogated. The primer extension analysis is then conducted on the purified PCR products. See PCT Patent Publications WO 91/13075; WO 92/15712; and WO 96/30545. Moreover, since considerable optimization is required to ensure that only the perfectly annealed oligonucleotide functions as a primer for the extension reaction, only limited multiplexing of the primer extension assays is possible. Multiplexing can be achieved by using primers of different lengths and by monitoring the wild-type and mutant nucleotide at each mutation site in two separate single nucleotide incorporation reactions. Such methods are provided herein.

Factors affecting hybridization are well known in the art and include temperature, ion concentration, pH, probe length, and probe GC content. A probe can hybridize at numerous places in an average genome. For example, any given 8-mer occurs about 65,000 times in the human genome. However, an octamer has a low melting temperature (Tm) and a single base mismatch will greatly exaggerate this instability. A 25-mer probe, for example, typically hybridizes with more stability than an 8-mer. However, because of the small thermodynamic differences in hybrid stability generated by single nucleotide changes, a longer probe will form a stable hybrid but will have a lower selectivity because it will tolerate nucleotide mismatches. Accordingly, under unfavorable hybridization conditions the short probe hybridizes with high selectivity (i.e., hybridizes poorly to sequence with even a single mismatch), but forms unstable hybrids. Longer probe will form a stable hybrid but will have a lower selectivity because of its tolerance of mismatches. To overcome this problem a mixture of probes is made of overlapping contiguous sequences of varying length. For example, the first and second probes hybridize to substantially contiguous portions of the target. For purposes of the present invention, substantially contiguous portions are those that are close enough together to allow hybridized first and second probes to function as a single probe. Substantially contiguous portions are preferably between zero (i.e., exactly contiguous so there is no space between the portions) nucleotides and about one nucleotide apart. It has now been realized that the adjacent probes bind cooperatively so that the longer, second probe imparts stability on the shorter, first probe. However, the stability imparted by the second probe does not overcome the selectivity (i.e., intolerance of mismatches) of the first probe. Therefore, methods of the invention take advantage of the high selectivity of the short first probe and the hybridization stability imparted by the longer second probe.

DNA molecules contain internucleotide phosphodiester linkages, which are degraded by exonucleases present in cells, culture media and human serum. For example, degradation by exonucleases in tissue culture media of DNA may be observed within about 30 minutes to about six hours. Various 3' exonucleases are known, such as phosphodiesterase from snake venom, exonuclease VII from E. coli, Bal 31 exonuclease, exonuclease III, 5' lambda exonuclease, and the 3'-5' exonuclease activity of some DNA polymerases exerted in the absence of dNTPs, as for example T4 DNA polymerase (See, e.g., U. S. Pat. No. 5,872,003 issued to Koster on February 16, 1999; U. S. Pat. No. 4,962,037 to Jett, et al., On October 9, 1990 and incorporated herein by reference). Methods are known to avoid this pitfall. For example, U.S.Pat. No. 5,256,775 to Froehler on October 26, 1993 discloses synthesis of exonuclease resistant nucleotides and as such this reference is incorporated herein by way of reference as means of preventing this eventuality. Another means of avoiding this inconvenience is to construct instead of DNA or RNA probes so-called peptide or polyamide nucleic acid (PNA) probes. PNA oligomers are synthesized like DNA oligomers in that the synthesis begins with the four bases (adenine, guanine, thymine, or cytosine) and linked together to form the oligomer of desired sequence. In the case of PNA, the monomers for making PNA each contain one of the four bases attached to 2-aminoethyl glycine. PNA monomers have amino and carboxyl termini, which are similar to amino acids. PNA monomers are thus linked by peptide bonds to form an oligomer and the synthesis protocols required to link the monomers are the same as those used for standard peptide synthesis. Non-limiting examples and details are found in U. S. Pat. No. 5,821,060 to Arlinghaus, et al., as incorporated herein by way of reference.

A variety of alternate methods have been developed which exploit sequence variation in DNA using enzymatic and chemical cleavage techniques. A commonly used screening method for DNA polymorphisms consists of digesting DNA with restriction endonucleases and analyzing the resulting fragments by means of Southern blots, as reported by Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1980 May;32(3):314-31 and incorporated herein by reference. RFLP analysis suffers from low sensitivity and requires a large amount of sample. When RFLP analysis is used for the detection of point mutations, it is, by its nature, limited to the detection of only those single base changes which fall within a restriction sequence of a known restriction endonuclease. Mutations that do not reside at the cleavage site of restriction endonuclease are not detected. One study reported that only 0.7% of the mutational variants were detected using RFLP analysis (Jeffreys AJ.DNA sequence variants in the G gamma-, A gamma-, delta- and beta-globin genes of man. Cell 1979 Sep;18(1):1-10).

It is an object of the present invention to provide a method for conducting affinity fractionation and sequencing of DNA comprising cleaving DNA into predetermined lengths; labeling the cleaved DNA; contacting the labeled DNA to an array of isolated oligonucleotide sequences, wherein said sequences are complementary to portions of the DNA; allowing hybridization to occur between said cleaved DNA and the sequences; determining the sequence of hybridized DNA from the sequences; contacting said extracted hybridized DNA. The technical details of this approach are well known to those skilled in art and are found, for example, in U. S. Pat. No. 5,905,024, as incorporated herein by way of reference.

The majority of the available enzymes have 4 to 6 base-pair recognition sequences, and cleave too frequently for many large-scale DNA manipulations (Eckstein and Lilley (eds.), Nucleic Acids and Molecular Biology, vol. 2, Springer-Verlag, Heidelberg, 1988). A handful of rare-cutting restriction enzymes with 8 base-pair specificities have been isolated and these are widely used in genetic mapping, but these enzymes are few in number, are limited to the recognition of G+C-rich sequences, and cleave at sites that tend to be highly clustered. Recently, endonucleases encoded by group I introns have been discovered that have greater than 12 base-pair specificity, see e.g., Perlman PS, Butow RA. Mobile introns and intron-encoded proteins. Science 1989 Dec 1;246(4934):1106-9, which incorporated herein by way of reference. Specific restriction enzymes or endonuclease useful for this approach are known in the art (see for example Molecular Cloning. A Laboratory Manual, 2nd Edition by Sambrook et al.) including but not limited to the group consisting of Aat II, gacgt/c, Acc 65I, g/gtacc, Acc I, gt/mkac, #Aci I, ccgc, Acl I, aa/cgtt, Afe I, agc/gct, Afl II, c/ttaag, Afl III, a/crygt, Age I, a/ccggt, Ahd I, gacnnn/nngtc, Alu I, ag/ct, #Alw I, ggatc, Alw NI, cagnnn/ctg, Apa I, gggcc/c, Apa LI, g/tgcac, Apo I, r/aatty, Asc I, gg/cgcgcc, Ase I, at/taat, Ava I, c/ycgrg, Ava II, g/gwcc, Avr II, c/ctagg, Bam HI, g/gatcc, Ban I, g/gyrcc, Ban II, grgcy/c, Bbe I, ggcgc/c, #Bbs I, gaagac, #Bbv CI, cctcagc, #Bbv I, gcagc, #Bcg I, gcannnnnntcg, #Bcg I, cgannnnnntgc, #Bci VI, gtatcc, Bcl I, t/gatca, Bfa I, c/tag, Bgl I, gccnnnn/nggc, Bgl II, a/gatct, Blp I, gc/tnagc, ;#Bmr I, actggg, Bpl I, gagnnnnnctcnnnnnnnnnnnnn/, Bpl I, gagnnnnnctcnnnnnnnnnnnnn/, #Bpm I, ctggag, #Bpu 10I, cctnagc, Bsa AI, yac/gtr, Bsa BI, gatnn/nnatc, Bsa HI, gr/cgyc, #Bsa I, ggtctc, Bsa JI, c/cnngg, Bsa WI, w/ccggw, #Bse MII, ctcag, #Bse RI, gaggag, #Bsg I, gtgcag, Bsi EI, cgry/cg, Bsi HKAI, gwgcw/c, Bsi WI, c/gtacg, Bsl I, ccnnnnn/nngg, #Bsm AI, gtctc, #Bsm BI, cgtctc, #Bsm FI, gggac, #Bsm I, gaatgc, Bso BI, c/ycgrg, Bsp 1286I, gdgch/c, Bsp DI, at/cgat, Bsp EI, t/ccgga, Bsp HI, t/catga, Bsp LU11I, a/catgt, #Bsp MI, acctgc, #Bsr BI, ccgctc, #Bsr DI, gcaatg, Bsr FI, r/ccggy, Bsr GI, t/gtaca, #Bsr I, actgg, Bss HII, g/cgcgc, Bss KI, /ccngg, #Bss SI, cacgag, Bst 4CI, acn/gt, Bst API, gcannnn/ntgc, Bst BI, tt/cgaa, Bst DSI, c/crygg, Bst EII, g/gtnacc, #Bst F5I, ggatg, Bst NI, cc/wgg, Bst UI, cg/cg, Bst XI, ccannnnn/ntgg, Bst YI, r/gatcy, Bst Z17I, gta/tac, Bsu 36I, cc/tnagg, #Btr I, cacgtc, Cac 8I, gcn/ngc, Cla I, at/cgat, Csp 6I, g/tac, Cvi JI, rg/cy, Dde I, c/tnag, Dpn I, ga/tc, Dpn II, /gatc, Dra I, ttt/aaa, Dra III, cacnnn/gtg, Drd I, gacnnnn/nngtc, Dsa I, c/crygg, Eae I, y/ggccr, Eag I, c/ggccg, #Ear I, ctcttc, Ecl 136II, gag/ctc, #Eco 57I, ctgaag, Eco ICRI, gag/ctc, Eco NI, cctnn/nnnagg, Eco O109I, rg/gnccy, Eco RI, g/aattc, Eco RII, /ccwgg, Eco RV, gat/atc, #Fau I, cccgc, Fnu 4HI, gc/ngc, #Fok I, ggatg, Fse I, ggccgg/cc, Fsp I, tgc/gca, Hae II, rgcgc/y, Hae III, gg/cc, #Hga I, gacgc, Hha I, gcg/c, Hin P1I, g/cgc, Hinc II, gty/rac, Hind III, a/agctt, Hinf I, g/antc, Hpa I, gtt/aac, Hpa II, c/cgg, #Hph I, ggtga, Kas I, g/gcgcc, Kpn I, ggtac/c, Mae II, a/cgt, Mae

III, /gtnac, Mbo I, /gatc, #Mbo II, gaaga, Mfe I, c/aattg, Mlu I, a/cgcgt, #Mnl I, cctc, Msc I, tgg/cca, Mse I, t/taa, Msl I, caynn/nnrtg, Msp A1I, cmg/ckg, Msp I, c/cgg, Mwo I, gcnnnnn/nngc, Nae I, gcc/ggc, Nar I, gg/cgcc, Nci I, cc/sgg, Nco I, c/catgg, Nde I, ca/tatg, Ngo MIV, g/ccggc, Nhe I, g/ctagc, Nla III, catg/, Nla IV, ggn/ncc, Not I, gc/ggccgc, Nru I, tcg/cga, Nsi I, atgca/t, Nsp I, rcatg/y, Pac I, ttaat/taa, Pae R7I, c/tcgag, Pci I, a/catgt, Pfl FI, gacn/nngtc, Pfl MI, ccannnn/ntgg, #Ple I, gagtc, Pme I, gttt/aaac, Pml I, cac/gtg, Ppu 10I, a/tgcat, Ppu MI, rg/gwccy, Psh AI, gacnn/nngtc, Psi I, tta/taa, Psp OMI, g/ggccc, Pst I, ctgca/g, Pvu I, cgat/cg, Pvu II, cag/ctg, Rsa I, gt/ac, Rsr II, cg/gwccg, Sac I, gagct/c, Sac II, ccgc/gg, Sal I, g/tcgac, San DI, gg/gwccc, #Sap I, gctcttc, Sau 3AI, /gatc, Sau 96I, g/gncc, Sbf I, cctgca/gg, Sca I, agt/act, #Sch I, gagtc, Scr FI, cc/ngg, Sex AI, a/ccwggt, #Sfa NI, gcatc, Sfc I, c/tryag, Sfi I, ggccnnnn/nggcc, Sfo I, ggc/gcc, Sgf I, gcgat/cgc, Sgr AI, cr/ccggyg, Sma I, ccc/ggg, Sml I, c/tyrag, Sna BI, tac/gta, Spe I, a/ctagt, Sph I, gcatg/c, Srf I, gccc/gggc, Ssp I, aat/att, Stu I, agg/cct, Sty I, c/cwwgg, Swa I, attt/aaat, Taa I, acn/gt, Tai I, acgt/, Taq I, t/cga, Tat I, w/gtacw, Tfi I, g/awtc, Tsc I, acgt/, Tse I, g/cwgc, Tsp 45I, /gtsac, Tsp 509I, /aatt, Tsp RI, castgnn/, Tth 111I, gacn/nngtc, Xba I, t/ctaga, Xcm I, ccannnnn/nnnntgg, Xho I, c/tcgag, Xma I, c/ccggg, Xmn I, gaann/nnttc, Xmn, gaann/nnttc or mixtures thereof (letters following enzyme name indicate the nucleotide sequence position at which these enzymes cleave the target nucleic acid; in addition to standard a, c, t, and g nucleotides additional letters represent ambiguous nucleotides such as r, which is equal to g or a; y=c or t; m=a or c; k=c or t; s=g or c; w=a or t; h=a or c or t; b=g or t or c; v=g or c or a; d=g or a or t; n=a or c or g or t). Other restricition enzymes are known and can be found for example in the U. S. Pat. No. 5, 861, 242 issued to Chee, et al., which is incorporated herein by way of reference.

5.3. Gene Expression and Quantitation Analysis Using Liquid Array

Gene expression and nucleic acid Quantitation analysis are as important as sequencing tasks by array technology. Multiple applications are familiar to one skilled in the art and are adaptable within the scope of the present invention.

Cytoplasmic RNA is extracted from cultured cells by the method of Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987 Apr;162(1):156-9, treated with DNAse I to remove DNA contamination, then extracted with phenol/chloroform and ethanol precipitated. Reverse transcriptions and PCR are performed to obtain cDNA as described in the exemplary differential display protocol disclosed by Nishio Y, Aiello LP, King GL. Glucose induced genes in bovine aortic smooth muscle cells identified by mRNA differential display. FASEB J 1994 Jan;8(1):103-6, as incorporated by way of reference. Prior to hybridization, PCR products are fluorescently labeled by random priming and unincorporated label is removed by Centricon filtration. Prior to hybridization, microparticles are rinsed with a solution of 1% "Blotto" or 50 mM tripolyphosphate and then washed with hybridization solution (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1M NaCl). Labeled PCR fragments representing the 3'-end of expressed genes are recovered from the Centricon filtration units in hybridization buffer and beads representing probes complementary to expressed genes of interest are flooded with appropriately diluted cDNA solution. The plate is placed at 65° C for 30-60 minutes and then washed three times with hybridization buffer. To facilitate washing and operating procedures the individual wells of microtiter plates have microporous filter bottoms to allow fluid passage but to retain microparticles with probes. Following hybridization and washing, the plate is placed into fluorescence reading plate reader and the intensity of signal is read. Other labeling means in addition to fluorescence detection are known, e.g., horse radish peroxidase, alkaline phosphatase, an isotope such as ^{32}P, chemiluminiscent reporter, and are disclosed for example in Eggers M, Hogan M, Reich RK, Lamture J, Ehrlich D, Hollis M, Kosicki B, Powdrill T, Beattie K, Smith S, et al. A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups. Biotechniques 1994 Sep;17(3):516-25, which is incorporated by way of reference. Many of these labeling methods are easily performed and kits are available from several commercial suppliers, e.g., Sigma, Pierce, Roche (non-covalent cis-platinum reagent), Vector Laboratories (labeling with linkers), GenHunter (random priming method), Enzo, Ambion (psoralen), Clontech, Vysis (nick translation method), PanVera (Cy3, Cy5, digoxin, rhodamine), Promega, Oncor, EpiCentre, Boehringer Mannheim, Molecular Probes (BODIPY, DNP, TAMRA, and fluorescein), Amersham Pharmacia Biotech (3'- and 5'-end labeling), Life Technologies, Schleicher and Schuell (intercalating agent labeling), etc. The choice is abundant and depending on the particular need one skilled in the art can select a suitable method (see for example, www.the-scientist.library.upenn.edu/yr1999/jan/profile1-990118.htm as incorporated herein by way of reference).

In order to have quantitative data the intensity of signal is compared to the signal from a reference material. Reference material is usually nucleic acid or molecule, which is identical or substantially homologous to the analyte of interest. The known quantity of reference material is diluted serially and standard curves are obtained accordingly, from which the quantity of the analyte is extrapolated.

Other methods of the analysis of gene expression are known and are easily adapted within the scope of this invention. This includes but not limited to serial analysis of gene expression (SAGE) method as disclosed in detail at www.sagenet.org the content of which is incorporated by way of reference.

5.4. TaqMan and Other Means of Identifying and Measuring Hybridization

TaqMan (synonymous with fluorogenic 5' nuclease assay) indicates the probe or Fluorescence Energy Transfer (FET) probe used to detect specific sequences in PCR products by employing the 5'>3' exonuclease activity of Taq DNA polymerase. The TaqMan probe (20-30 bp), disabled from extension at the 3' end, consists of a site-specific sequence labeled with a fluorescent reporter dye and a fluorescent quencher dye. During PCR the TaqMan probe hybridizes to its complementary single strand DNA sequence within the PCR target. When amplification occurs the TaqMan probe is degraded due to the 5'-->3' exonuclease activity of Taq DNA polymerase, thereby separating the quencher from the reporter during extension. Due to the release of the quenching effect on the reporter, the fluorescence intensity of the reporter dye increases. This allows the real time (kinetic) quantitation of PCR products using a target-specific, energy transfer probe. During the entire amplification process this light emission increases exponentially, the final level being measured by spectrophotometry after termination of the PCR. Because increase of the fluorescence intensity of the reporter dye is only achieved when probe hybridization and amplification of the target sequence has occurred, the TaqMan assay offers a sensitive method to determine the presence or absence of specific sequences. Therefore, this technique is particularly useful in diagnostic applications, such as the screening of samples for the presence or incorporation of favorable traits and the detection of pathogens and diseases. The TaqMan assay allows high sample throughput because no gel-electrophoresis is required for detection. When different probes are used which are able to discriminate between allelic variants, TaqMan behaves as a codominant marker. The technical details regarding TaqMan assay are found in Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A 1991 Aug 15;88(16):7276-80; Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome Res 1996 Oct;6(10):986-94; and also at the web site by Molecular Beacons (http://molecular-beacons.org), all these references are incorporated herein by reference.

A method similar to Taqman approach is based on use of PNAs, which are polyamide analogs of DNA and RNA. See, e.g., U. S. Pat. No. 5,539,082 to Nielsen et al. Nielsen et al. discloses that PNAs simulate natural polynucleotides by binding complementary single stranded (ss) DNA and RNA strands. The method is conducted without separating unhybridized probes from the hybridization complex prior to signal detecting, without providing a signal quenching agent on the probe, or on the nucleotide sequence of interest, and without the use of enzymes. The details of this method are provided in U. S. Pat. No. 5,846,729 issued to Wu, et al. December 8, 1998 and incorporated herein by way of reference.

In addition to Taqman and PNA other alternative labeling or tagging methods are known in the art, e.g., U. S. Pat. Nos. 5,869,245 to Yeung on February 9, 1999; 5,759,779 to Dehlinger on June 2, 1998 as incorporated by reference and contemplated within the scope of the invention.

For direct and indirect labeling methods the samples of nucleic acid are labeled with fluorescein and biotin, respectively. Reactions are carried out at 1.25 mM of ATP, CTP, GTP, and UTP and 0.5 mM fluorescein-12-UTP or 0.25 mM biotin-16-UTP (Boehringer Mannheim). Labeled samples are denatured at 95 degrees C for 5 min, chilled on ice for 5 min, and equilibrated to 37 degrees C. See for further details in Melchior WB Jr, Von Hippel PH. Alteration of the relative stability of dA-dT and dG-dC base pairs in DNA. Proc Natl Acad Sci U S A 1973 Feb;70(2):298-302; Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP. Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 1995 Sep;19(3):442-7, both of which are incorporated by way of reference. A hybridization solution is then added to each reaction vessel of microtiter plate and microparticles are incubated with samples at 37 degrees C for 15 minutes. The particles are washed with 6 x SSPE (0.9 M NaCl, 60 mM Na H.sub.2 PO.sub.4 , 6 mM EDTA, pH 7.4 with 0.005% Triton X-100). Phycoerythrin-conjugated streptavidin (2 μg/ml in 6x SSPE, 0.005% Triton X-100) is added to bitin incorporating DNA samples and incubation continued at room temperature for 5 min. The microparticles are washed and scanned at a resolution of 74 pixels in a fluorescence plate reader. Two sets of data are collected: a fluorescein scan is obtained with a 515- to 545-nm band-pass fitter and a phycoerythrin scan with a 560-nm long-pass filter.

5.5. Nucleic Acid Amplification

In many applications, e.g., detection of mutations, the target nucleic acid sequence needs to be amplified. A handful of methods have been devised to amplify nucleic acid segments. This approach avoids the time and expense associated with cloning the segment of interest. The Polymerase Chain Reaction (PCR) is the original method for nucleic acid amplification. However, several other methods have been developed that employ the same basis of specificity, but create signal by different amplification mechanisms. These methods include but are not limited to the Ligase Chain Reaction (LCR), Self-Sustained Synthetic Reaction (3SR/NASBA), and Q-beta Replicase.

The polymerase chain reaction (PCR), as disclosed in U.S. Pat. Nos. 4, 683, 195 and 4, 683, 202 to Mullis and Mullis et al., is a method for increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. This technology provides one approach to the problems of low target sequence concentration. PCR is usually used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves introducing an excess of two oligonucleotide primers, which are complementary to their respective strands of the double-stranded target sequence in the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of the denaturation, hybridization, and polymerase extension are repeated as often as needed, in order to obtain high concentrations of a segment of the desired target sequence. The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified."

The ligase chain reaction (LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR) described by Wiedmann M, Wilson WJ, Czajka J, Luo J, Barany F, Batt CA. Ligase chain reaction (LCR)--overview and applications. PCR Methods Appl 1994 Feb;3(4):S51-64, as incorporated herein by reference, has developed into a well-recognized alternative method for amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes. Segev, PCT Publication No. W09001069 A1 (1990). However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

The self-sustained sequence replication reaction (3SR/NASBA) (Guatelli JC, Whitfield KM, Kwoh DY, Barringer KJ, Richman DD, Gingeras TR. Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. Proc Natl Acad Sci U S A 1990 Mar;87(5):1874-8) is a transcription-based in vitro amplification system that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA is then utilized for mutation detection (Fahy E, Kwoh DY, Gingeras TR. Self-sustained sequence replication (3SR): an isothermal transcription-based amplification system alternative to PCR. PCR Methods Appl 1991 Aug;1(1):25-33). In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo-and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (e.g., 200-300 base pairs).

Q-Beta Replicase. In this method, a probe which recognizes the sequence of interest is attached to the replicatable RNA template for Q-beta replicase, e.g., see incorporated by reference an article by Abramson RD, Myers TW. Nucleic acid amplification technologies. Curr Opin Biotechnol 1993 Feb;4(1):41-7. A previously identified major problem with false positives resulting from the replication of unhybridized probes has been addressed through use of a sequence-specific ligation step. However, available thermostable DNA ligases are not effective on this RNA substrate, so the ligation must be performed by T4 DNA ligase at low temperatures. This prevents the use of high temperature as a means of achieving specificity as in the LCR, the ligation event can be used to detect a mutation at the junction site, but not elsewhere.

The cycling probe reaction (CPR) uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the remainder of the probe from the target DNA and allowing another probe molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may carried through sample preparation.

Branched DNA (bDNA), described by Urdea et al., (Urdea MS, Running JA, Horn T, Clyne J, Ku LL, Warner BD. A novel method for the rapid detection of specific nucleotide sequences in crude biological samples without blotting or radioactivity; application to the analysis of hepatitis B virus in human serum. Gene 1987;61(3):253-64), involves oligonucleotides with branched structures that allow each individual oligonucleotide to carry 35 to 40 labels (e.g., alkaline phosphatase enzymes). The details of the technique are found in U. S. Pat. No. 5,597,909 to Urdea, et al. Issued on January 28, 1997. While this enhances the signal from a hybridization event, signal from non-specific binding is similarly increased.

5.6. SNP

Methods have been devised to detect the presence or absence of mutations within disease-associated genes. One such method is to compare the complete nucleotide sequence of a sample genomic region with the corresponding wild-type region. See, for example, Engelke DR, Hoener PA, Collins FS. Direct sequencing of enzymatically amplified human genomic DNA. Proc Natl Acad Sci U S A 1988 Jan;85(2):544-8 and Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazian HH Jr. Characterization of beta-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. Nature 1987 Nov 26-Dec 2;330(6146):384-6. However, such methods are costly, time consuming, and require the extensive cloning and sequencing for unambiguous detection of low-frequency mutations. As such, this method it is not practical to use for routine screening of genetic mutations.

Single nucleotide polymorphisms (SNPs) are DNA point mutations and is estimated that SNPs occur once every 100-300 base pair (bp). SNPs are useful for gene mapping, defining population structure, and performing functional studies. SNPs are expected to greatly facilitate large-scale genetic studies concerned with determining linkage between sequence variations and heritable phenotypes. SNPs are also an efficient tool for genetic identification for legal and forensic applications. SNPs are particularly interesting as markers because many known genetic diseases, such as sickle cell anemia, are caused by single base mutations. Therefore, an assay for an SNP marker is useful for identification of the disease-causing mutation. Some genetic diseases, such as cystic fibrosis, are caused by any of a large number of different mutations in a single gene. In this situation SNPs are used in a multiplex assay for all known alleles in a large gene. SNPs have been proposed as an ideal tool for the emerging discipline of pharmacogenomics for fine-tuning of patients' diagnosis and treatment. Large-scale analysis of the associations between the effects of drugs and genetic markers allows physicians to match drugs to the genetic makeup of individual patients to better predict beneficial and harmful effects. Databases of gene expression profiles can be predictive of different classes of drug toxicity.

Some advantages of SNPs over other types of genetic markers such as isozymes, restriction fragment length polymorphisms (RFLP), variable number tandem repeats and simple sequence repeats include: very large numbers of polymorphic loci; loci distributed throughout the genome; markers present within coding regions or exons, introns and regions that flank exons; simple and unambiguous assay techniques; high levels of polymorphism in the population; stable Mendelian inheritance; and low levels of spontaneous mutation. Kwok et. al (Kwok, P.Y., Q. Deng, H. Zakeri, S.L. Taylor and D.A. Nickerson. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSS. Genomics 31:123-126, 1996) have observed that most of the SNPs identified from sequence databases are highly heterozygous in the population.

5.7. SNPs on the Web

Most of data on currently known SNPs are available publicly in databases on the World Wide Web (www), a global computer network, and one skilled in the art easily finds them by selecting appropriate key words in a search engine (Gu Z, Hillier L, Kwok PY. Single nucleotide polymorphism hunting in cyberspace. Hum Mutat 1998;12(4):221-5; Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M. Mining SNPs from EST databases. Genome Res 1999 Feb;9(2):167-74; and Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. Genome Res 1998 Jul;8(7):748-54). The content of these databases is disclosed hereinafter and incorporated in their entirety in this specification. In September 1998, the National Human Genome Research Institute (NHGRI) and National Center for Biotechnology Information (NCBI established a database of SNPs on (www.ncbi.nlm.nih.gov/SNP) to serve as a central repository of database (dbSNP) sequences. This database contains sequence information flanking each known polymorphism, links to related GenBank® accessions and to the GenBank database of STS (dbSTS). The NHGRI and the Centers for Disease Control (CDC) have developed a set of "standard" human DNA samples (NIH Polymorphism Discovery Resource (NIHPDR) for evaluation of polymorphic markers. This resource will comprise cell lines and DNA from 450 unrelated individuals, female and male. The material in the resource is available to any investigator from the Coriell Institute for Medical Research (Camden, NJ, USA) at http://arginine.umdnj.edu/. The National Cancer Institute (NCI) project, know as the Genetic Annotation Initiative (GAI), looks for SNPs by sequencing the coding regions and flanking regions of several thousand genes suspected of contributing to cancer susceptibility or resistance. SNPs discovered by the GAI are added to the NCBI's public dbSNP. This complements the NCI's Cancer Genome Anatomy project, which collects STS data from a broad set of cancerous and noncancerous tissues. The Whitehead Institute/MIT Center for Genome Research, has a Human SNP Mapping Project funded by a consortium consisting of Affymetrix, Bristol-Myers Squibb and Millennium Pharmaceuticals and data are available on www.genome.wi.mit.edu/SNP/. Washington University at St. Louis has its own SNP database on (www.ibc.wustl.edu/SNP). This database is organized by cytogentic location (chromosome arm). Each SNP has a polymerase chain reaction (PCR) primer and assay conditions associated with it. Web site from Johns Hopkins University has also a genetic database found at www.bis.med.jhmi.edu/. A similar site from Weizmann Institute is found at http://bioinformatics.weizmann.ac.il/cards/. Human mitochondrial sequence is found at web site of Emory University: http://infinity.gen.emory.edu/mitomap.html. A large number of SNPs have already been identified using public data in GenBank from the genome sequencing projects and expressed sequence tag (EST) sequences. SNPs are found by scanning for regions of overlapping sequence in bacterial artificial chromosome (BAC) and PI-derived artificial chromosome (PAC) clones that come from different individuals or from homologous chromosomes in a single individual. It is also possible to discover SNPs by scanning for variation among homologous EST sequences. The EST approach is particularly attractive since these sequences are by definition derived from expressed genes, and in fact, some SNPs may represent mutations with detectable phenotypes. An additional advantage of this approach is that many of these database sequences have already been mapped. Several private companies are compiling their own sets of SNPs by large-scale sequencing of identical genomic regions from genetically diverse individuals. CuraGen (New Haven, CT, USA) collected over 60,000 SNPs located in expressed genes (www.curagen.com/). Celera Genomics (Rockville, MD, USA; www.celera.com) (a collaboration between Perkin-Elmer and The Institute for Genomic Research) and Incyte Pharmaceuticals (Palo Alto, CA, USA; www.incyte.com) have private databases of approximately 100,000 SNPs. Other companies involved in developing SNP databases include Genset (Paris, France; www.genxy.com/Science), Orchid Biocomputer (Princeton, NJ, USA; www.orchidbio.com), Nanogen (vww.nanogen.com/), Eos (www.eosbiotech.com), Affymetrix (www.affymetrix.com), and Variagenics (Cambridge, MA, USA). The entire content of these web sites and updates thereof is provided and incorporated by way of reference and specific examples listed hereinabove are not in any way limiting.

The instant invention is particularly well suited to SNP assays, since large numbers of SNPs are assayed simultaneously on a single array.

6. EXAMPLES

The preparation of multiple subsets of microspheres having unique fluorescence signatures is described elsewhere, for example, in International Patent Application (PCT) No. WO99/19515, published April 22, 1999, the complete disclosure of which is incorporated by reference herein.

Many practical applications of the instant invention are readily recognizable to those of ordinary skill in the art. Applications involving sequencing by hybridization, detection of aberrant or normal gene expression, and quantitation of nucleic acid of interest are non-limiting three main applications that are readily supported by this invention. At the present time there are about 4,000 recognized genetic diseases in humans. This invention encompasses them and those to be discovered in the future and is not limited to those listed hereinafter as specific examples of strategies of how to deal with identification and analysis of mutations versus non-mutations. The publications as disclosed infra are expressly incorporated in their entirety by way of reference.

6.1. Applications in a Variety of Malignant Diseases

In 1994, two breast cancer susceptibility genes were identified: BRCA1 on chromosome 17 and BRCA2 on chromosome 13. When an individual carries a mutation in either BRCA1 or BRCA2, they are at an increased risk of being diagnosed with breast or ovarian cancer at some point in their lives. Until recently, it was not clear what the function of these genes was, until studies on a related protein in yeast revealed their normal role: they participate in repairing radiation-induced breaks in double-stranded DNA. It is though that mutations in BRCA1 or BRCA2 disable this mechanism, leading to more errors in DNA replication and ultimately to cancerous growth.

Oncogene myc mutation is associated with Burkitt lymphoma (Bhatia K, Huppi K, Spangler G, Siwarski D, Iyer R, Magrath I. Point mutations in the c-Myc transactivation domain are common in Burkitt's lymphoma and mouse plasmacytomas. Nat Genet 1993 Sep;5(1):56-61). Mutations in MSH2 and MLH1 are associated with colon cancer (Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994 Mar 17;368(6468):258-61; Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993 Dec 3;75(5):1027-38).

Lung cancer is also associated with genetic mutations (Braun MM, Caporaso NE, Page WF, Hoover RN. Genetic component of lung cancer: cohort study of twins. Lancet 1994 Aug 13;344(8920):440-3; Sjalander A, Birgander R, Rannug A, Alexandrie AK, Tornling G, Beckman G. Association between the p21 codon 31 A1 (arg) allele and lung cancer. Hum Hered 1996 Jul-Aug;46(4):221-5; and Weston A, Willey JC, Modali R, Sugimura H, McDowell EM, Resau J, Light B, Haugen A, Mann DL, Trump BF, et al. Differential DNA sequence deletions from chromosomes 3, 11, 13, and 17 in squamous-cell carcinoma, large-cell carcinoma, and adenocarcinoma of the human lung. Proc Natl Acad Sci U S A 1989 Jul;86(13):5099-103).

Malignant melanoma exhibits genetic heterogeneity and methods of detecting them are disclosed by Hussussian et al., (Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH Jr, Tucker MA, Dracopoli NC. Germline p16 mutations in familial melanoma. Nat Genet 1994 Sep;8(1):15-21).

Multiple endocrine neoplasia (MEN) is a group of rare diseases caused by genetic defects that lead to hyperplasia (abnormal multiplication or increase in the number of normal cells in normal arrangement in a tissue) and hyperfunction (excessive functioning) of 2 or more components of the endocrine system. Genetic mutations associated with MEN are known (Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SE, Collins FS, Emmert-Buck MR, Debelenko LV, Zhuang Z, Lubensky IA, Liotta LA, Crabtree JS, Wang Y, Roe BA, Weisemann J, Boguski MS, Agarwal SK, Kester MB, Kim YS, Heppner C, Dong Q, Spiegel AM, Burns AL, Marx SJ. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science 1997 Apr 18;276(5311):404-7).

Neurofibramatosis type 2 (NF-2) is a rare inherited disorder characterized by the development of benign tumors on both auditory nerves (acoustic neuromas). The disease is also characterized by the development of malignant central nervous system tumors as well. The NF2 gene has been mapped to chromosome 22 and is thought to be a so-called 'tumor- suppressor gene'. Like other tumor suppressor genes (such as p53 and Rb), the normal function of NF2 is to act as a brake on cell growth and division, ensuring that cells do not divide uncontrollably, as they do in tumors. A mutation in NF2 impairs its function, and accounts for the clinical symptoms observed in neurofibromatosis sufferers (Rouleau GA, Merel P, Lutchman M, Sanson M, Zucman J, Marineau C, Hoang-Xuan K, Demczuk S, Desmaze C, Plougastel B, et al. Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature 1993 Jun 10;363(6429):515-21).

The p53 gene like the Rb gene, is a tumor suppressor gene, i.e., its activity stops the formation of tumors. If a person inherits only one functional copy of the p53 gene from their parents, they are predisposed to cancer and usually develop several independent tumors in a variety of tissues in early adulthood. This condition is rare, and is known as Li-Fraumeni syndrome. However, mutations in p53 are found in most tumor types, and so contribute to the complex network of molecular events leading to tumor formation. The p53 gene has been mapped to chromosome 17. In the cell, p53 protein binds DNA, which in turn stimulates another gene to produce a protein called p21 that interacts with a cell division-stimulating protein (cdk2). When p21 is complexed with cdk2 the cell cannot pass through to the next stage of cell division. Mutant p53 no longer binds DNA in an effective way, and as a consequence the p21 protein is not made available to act as the 'stop signal' for cell division. Thus cells divide uncontrollably, and form tumors (Harlow E, Williamson NM, Ralston R, Helfman DM, Adams

TE. Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53. Mol Cell Biol 1985 Jul;5(7):1601-10; Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994 Jul 15;265(5170):346-55). Other oncogenes or tumor suppressor genes of a eukaryotic (for example, mammalian) cell are contemplated as useful markers for malignancies; preferable mammalian oncogenes include, without limitation, abl, akt, crk, erb-A, erb-B, ets, fes/fps, fgr, fms, fos, jun, kit, mil/raf, mos, myb, myc, H-ras, K-ras, rel, ros, sea, sis, ski, src, and yes; preferable tumor suppressor genes in addition to p53 include, retinoblastoma (preferably RB1), adenomatous polyposis coli, NF-1, NF-2, MLH-1, MTS-1, MSH-2, and human non-polyposis genes. Reference nucleic acids are also derived from any cell cycle control gene, preferably p21, p27, or p16. (U. S. Pat. No. 5,876,941 to Landegren, et al., on March 2, 1999).

About 90% of human pancreatic carcinomas show a loss of part of chromosome 18. In 1996, a possible tumor suppressor gene, DPC4 (Smad4), was discovered from the section that is lost in pancreatic cancer. There is a whole family of Smad proteins in vertebrates, all involved in signal transduction of transforming growth factor-beta (TGF-beta) related pathways. Other tumor suppressor genes include p53 and Rb, which, if mutated or absent from the genome can contribute to cancerous growth in a variety of tissues (Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 1996 Jan 19;271(5247):350-3).

Despite the high prevalence of prostate cancer, little is known about the genetic predisposition of some men to the disease. Numerous studies point to a family history being a major risk factor and thought to be responsible for an estimated 5-10% of all prostate cancers. The discovery of a susceptibility locus for prostate cancer on chromosome 1, called HPC1, accounts for about 1 in 500 cases of prostate cancer (Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstein MJ, Bova GS, Guo H, Bujnovszky P, Nusskern DR, Damber JE, Bergh A, Emanuelsson M, Kallioniemi OP, Walker-Daniels J, Bailey-Wilson JE, Beaty TH, Meyers DA, Walsh PC, Collins FS, Trent JM, Isaacs WB. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996 Nov 22;274(5291):1371-4).

While no two cancers are genetically identical (even in the same tissue type), there are relatively few ways in which normal cell growth can go wrong. One of these is to make a gene that stimulates cell growth hyperactive; this altered gene is known as an oncogene. Ras is one such oncogene product that is found on chromosome 11. It is found in normal cells, where it helps to relay signals by acting as a switch. When receptors on the cell surface are stimulated (by a hormone, for example), Ras is switched on and transduces signals that tell the cell to grow. If the cell-surface receptor is not stimulated, Ras is not activated and so the pathway that results in cell growth is not initiated. In about 30% of human cancers, Ras is mutated (Lowy DR, Willumsen BM. Function and regulation of ras. Annu Rev Biochem 1993;62:851-91; Russell MW, Munroe DJ, Bric E, Housman DE, Dietz-Band J, Riethman HC, Collins FS, Brody LC. A 500-kb physical map and contig from the Harvey ras-1 gene to the 11p telomere. Genomics 1996 Jul 15;35(2):353-60; Tong LA, de Vos AM, Milburn MV, Jancarik J, Noguchi S, Nishimura S, Miura K, Ohtsuka E, Kim SH. Structural differences between a ras oncogene protein and the normal protein. Nature 1989 Jan 5;337(6202):90-3).

Retinoblastoma occurs in early childhood and affects about 1 child in 20,000. There are both hereditary and non-hereditary forms of retinoblastoma. In the hereditary form, multiple tumors are found in both eyes, while in the non-hereditary form only one eye is effected and by only one tumor. In the hereditary form, a gene called Rb is lost from chromosome 13. Since the absence of Rb seemed to be linked to retinoblastoma, it has been suggested that the role of Rb in normal cells is to suppress tumor formation. Rb is found in all cells of the body, where under normal conditions it acts as a brake on the cell division cycle by preventing certain regulatory proteins from triggering DNA replication. If Rb is missing, a cell can replicate itself over and over in an uncontrolled manner, resulting in tumor formation. Untreated, retinoblastoma is almost uniformly fatal, but with early diagnosis and modern methods of treatment the survival rate is over 90%. Since the Rb gene is found in all cell types, studying the molecular mechanism of tumor suppression by Rb will give insight into the progression of many types of cancer, not just retinoblastoma (Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 1986 Oct 16-22;323(6089):643-6; Lee WH, Bookstein R, Hong F, Young LJ, Shew JY, Lee EY. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 1987 Mar 13;235(4794):1394-9).

Von Hippel-Lindau Syndrome is an inherited multi-system disorder characterized by abnormal growth of blood vessels. While blood vessels normally grow like trees, in people with VHL little knots of blood capillaries sometimes occur. These knots are called angiomas or hemangioblastomas. Growths may develop in the retina, certain areas of the brain, the spinal cord, the adrenal glands and other parts of the body. The gene for Von-Hippel Lindau disease (VHL) is found on chromosome 3, and is inherited in a dominant fashion. If one parent has a dominant gene, each child has a 50-50 chance of inheriting that gene. The VHL gene is a tumor suppressor gene. This means that its role in a normal cell is to stop uncontrolled growth and proliferation. If the gene is lost or mutated, then its inhibitory effect on cell growth is lost or diminished, which, in combination with defects in other regulatory proteins, can lead to cancerous growth (Latif F, Tory K, Gnarra J, Yao M, Duh FM, Orcutt ML, Stackhouse T, Kuzmin I, Modi W, Geil L, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. Science 1993 May 28;260(5112):1317-20).

6.2. Application for Detection and Treatment of Immune Disorders

Asthma is a what is known as a complex heritable disease. This means that there are a number of genes that contribute towards a person's susceptibility to a disease, and in the case of asthma, chromosomes 5, 6, 11, 14, and 12 have all been implicated. The relative roles of these genes in asthma predisposition are not clear, but one of the most promising sites for investigation is on chromosome 5. Although a gene for asthma from this site has not yet been specifically identified, it is known that this region is rich in genes coding for key molecules in the inflammatory response seen in asthma, including cytokines, growth factors, and growth factor receptors. The search for specific asthma genes can be assisted by the present invention (A genome-wide search for asthma susceptibility loci in ethnically diverse populations. The Collaborative Study on the Genetics of Asthma (CSGA). Nat Genet 1997 Apr;15(4):389-92).

Autoimmune polyglandular syndrome type I (APS1, also called APECED) is a rare autosomal recessive disorder that maps to human chromosome 21. At the end of 1997, researchers reported that they isolated a novel gene, which they called AIRE (autoimmune regulator). Database searches revealed that the protein product of this gene is a transcription factor - a protein that plays a role in the regulation of gene expression. The researchers showed that mutations in this gene are responsible for the pathogenesis of APS1 (Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, Krohn KJ, Lalioti MD, Mullis PE, Antonarakis SE, Kawasaki K, Asakawa S, Ito F, Shimizu N. Positional cloning of the APECED gene. Nat Genet 1997 Dec;17(4):393-8).

Inflammatory bowel diseases (IBD) is a group of chronic disorders that cause inflammation or ulceration in the small and large intestines. Often, IBD is classified either as ulcerative colitis or Crohn disease. A susceptibility locus for the disease was recently mapped to chromosome 16. Candidate genes found in this region include several involved in the inflamatory response, including: CD19, involved in B-lymphocyte function; sialophorin, involved in leukocyte adhesion; the CD11 integrin cluster, involved in microbacterial cell adhesion; and the interleukin-4 receptor, which is interesting, as IL-4-mediated functions are altered in IBDs (Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996 Feb 29;379(6568):821-3).

DiGeorge syndrome is a rare congenital (i.e. present at birth) disease whose symptoms vary greatly between individuals, but commonly include a history of recurrent infection, heart defects and characteristic facial features. DiGeorge syndrome is caused by a large deletion from chromosome 22, produced by an error in recombination at meiosis (the process that creates germ cells and ensures genetic variation in the offspring). This deletion means that several genes from this region are not present in DiGeorge syndrome patients. It appears that the variation in the symptoms of the disease is related to the amount of genetic material lost in the chromosomal deletion (Demczuk S, Thomas G, Aurias A. Isolation of a novel gene from the DiGeorge syndrome critical region with homology to Drosophila gdl and to human LAMC1 genes. Hum Mol Genet 1996 May;5(5):633-8; Budarf ML, Collins J, Gong W, Roe B, Wang Z, Bailey LC, Sellinger B, Michaud D, Driscoll DA, Emanuel BS. Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. Nat Genet 1995 Jul;10(3):269-78).

Familial Mediterranean fewer (FMF) is an inherited disorder usually characterized by recurrent episodes of fever and peritonitis (inflammation of the abdominal membrane). In 1997, researchers identified the gene for FMF and found several different gene mutations that cause this inherited rheumatic disease. The gene, found on chromosome 16, codes for a protein that is found almost exclusively in granulocytes. The protein, pyrin, is a member of a family of nuclear factors homologous to the Ro52 autoantigen and is likely to normally assist in keeping inflammation under control. Discovery of the gene mutations will allow the development of a simple diagnostic blood test for FMF (Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. The International FMF Consortium. Cell 1997 Aug 22;90(4):797-807; A candidate gene for familial Mediterranean fever. The French FMF Consortium. Nat Genet 1997 Sep;17(1):25-31).

Severe combined immunodeficiency (SCID) also known as "bubble boy" disease represents a group of rare, sometimes fatal, congenital disorders characterized by little or no immune response. All forms of SCID are inherited, with as many as half of SCID cases linked to the X chromosome, passed on by the mother. X-linked SCID results from a mutation in the interleukin 2 receptor gamma (IL2RG) gene which produces the common gamma chain subunit, a component of several IL receptors. In another form of SCID, there is a lack of the enzyme adenosine deaminase (ADA), coded for by a gene on chromosome 20. This means that the substrates for this enzyme accumulate in cells. Immature lymphoid cells of the immune system are particularly sensitive to the toxic effects of these unused substrates, so fail to reach maturity (Valerio D, Duyvesteyn MG, Dekker BM, Weeda G, Berkvens TM, van der Voorn L, van Ormondt H, van der Eb AJ. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J 1985 Feb;4(2):437-43; Noguchi M, Yi H, Rosenblatt HM, Filipovich AH, Adelstein S, Modi WS, McBride OW, Leonard WJ. Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. Cell 1993 Apr 9;73(1):147-57).

Susceptibility to several disorders, including insulin-dependent diabetes mellitus and multiple sclerosis, is associated with alleles of HLA class II genes, i.e., IKBL gene, which lies near the TNF cluster at the telomeric end of the central major histocompatibility complex (MHC) region (Allcock RJ, Christiansen FT, Price P. The central MHC gene IKBL carries a structural polymorphism that is associated with HLA-A3,B7,DR15. Immunogenetics 1999 Jun 8;49(7/8):660-665).

6.3. Diagnosis and Therapy of Muscle and Bone Diseases

Duchenne muscular dystrophy (DMD) is one of a group of muscular dystrophies characterized by the enlargement of muscles. DMD is one of the most prevalent types of muscular dystrophy and is characterized by rapid progression of muscle degeneration which occurs early in life. All are X-linked and affect mainly males - an estimated 3,500 boys worldwide. The gene for DMD, found on the X chromosome, encodes a large protein - dystrophin. Dystrophin is required inside muscle cells for structural support: it is thought to strengthen muscle cells by anchoring elements of the internal cytoskeleton to the surface membrane. Without it, the cell membrane becomes permeable, so that extracellular components enter the cell, increasing the internal pressure until the muscle cell dies (Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 1988 Apr 22;53(2):219-26).

Ellis-van Creveld syndrome, also known as 'chondroectodermal dysplasia', is a rare genetic disorder characterized by short-limb dwarfism, polydactyly (additional fingers or toes), malformation of the bones of the wrist, dystrophy of the fingernails, partial hare-lip, cardiac malformation and often prenatal eruption of the teeth. The gene causing Ellis-van Creveld syndrome, EVC, has been mapped to the short arm of chromosome 4. As yet, the function of a healthy EVC gene is not known (Polymeropoulos MH, Ide SE, Wright M, Goodship J, Weissenbach J, Pyeritz RE, Da Silva EO, Ortiz De Luna RI, Francomano CA. The gene for the Ellis-van Creveld syndrome is located on chromosome 4p16. Genomics 1996 Jul 1;35(1):1-5).

Marfan syndrome is a connective tissue disorder, so affects many structures, including the skeleton, lungs, eyes, heart and blood vessels. The disease is characterized by unusually long limbs. Marfan syndrome is an autosomal dominant disorder that has been linked to the FBN1 gene on chromosome 15. FBN1 encodes a protein called fibrillin, which is essential for the formation of elastic fibres found in connective tissue (Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, et al. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 1991 Jul 25;352(6333):337-9).

Myotonic dystrophy is an inherited disorder in which the muscles contract but have decreasing power to relax. With this condition, the muscles also become weak and waste away. Myotonic dystrophy can cause mental deficiency, hair loss and cataracts. The myotonic dystrophy gene, found on chromosome 19, codes for a protein kinase that is found in skeletal muscle, where it likely plays a regulatory role. An unusual feature of this illness is that its symptoms usually become more severe with each successive generation. This is because mistakes in the faithful copying of the gene from one generation to the next result in the amplification of a 'AGC triplet repeat', similar to that found in Huntington disease. Unaffected individuals have between 5 and 27 copies of AGC, myotonic dystrophy patients who are minimally affected have at least 50 repeats, while more severely affected patients have an expansion of up to several kilobase pairs (Aslanidis C, Jansen G, Amemiya C, Shutler

G, Mahadevan M, Tsilfidis C, Chen C, Alleman J, Wormskamp NG, Vooijs M, et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature 1992 Feb 6;355(6360):548-51).

6.4. Application in Diseases of Nervous System

Alzheimer disease (AD) is the fourth leading cause of death in adults. The incidence of the disease rises steeply with age. AD is twice as common in women than in men. Some of the most frequently observed symptoms of the disease include a progressive inability to remember facts and events and, later, to recognize friends and family. AD tends to run in families: currently, mutations in four genes, situated on chromosomes 1, 14, 19 and 21, are believed to play a role in the disease. The best-characterized of these are PS1 (or AD3) on chromosome 14 and PS2 (or AD4) on chromosome 1. The formation of lesions made of fragmented brain cells surrounded by amyloid-family proteins are characteristic of the disease. Interestingly, these lesions and their associated proteins are closely related to similar structures found in Down's Syndrome (Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 1995 Aug 18;269(5226):973-7; Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 1995 Jun 29;375(6534):754-60).

Amyotrophic lateral sclerosis (ALS) or Lou Gehrig disease is a neurological disorder characterized by progressive degeneration of motor neuron cells in the spinal cord and brain, which ultimately results in paralysis and death. In 1991, a team of researchers linked familial ALS to chromosome 21. Two years later, the SOD1 gene was identified as being associated with many cases of familial ALS. The enzyme coded for by SOD1 carries out a very important function in cells: it removes dangerous superoxide radicals by converting them into non-harmful substances (Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 1993 Mar 4;362(6415):59-62).

Charcot-Marie-Tooth disease (CMT) disease is the most common inherited peripheral neuropathy in the world, characterized by a slowly progressive degeneration of the muscles in the foot, lower leg, hand and forearm, and a mild loss of sensation in the limbs, fingers and toes. Full expression of CMT's clinical symptoms generally occurs by age 30. CMT is not a fatal disease, however, and the disorder does not affect normal life expectancy. CMT is a genetically heterogeneous disorder, in which mutations in different genes can produce the same clinical symptoms. In CMT, there are not only different genes but different patterns of inheritance. One of the most common forms of CMT is Type 1A. The gene for Type 1A CMT maps to chromosome 17 and is thought to code for a protein (PMP22) involved in coating peripheral nerves with myelin, a fatty sheath that is important for their conductance. Other types of CMT include Type 1B, autosomal-recessive and X-linked. The same proteins involved in the Type 1A and Type 1B CMT are also involved in a disease called Dejerine-Sottas syndrome (DSS), in which similar clinical symptoms are presented, but they are more severe (Hayasaka K, Himoro M, Wang Y, Takata M, Minoshima S, Shimizu N, Miura M, Uyemura K, Takada G. Structure and chromosomal localization of the gene encoding the human myelin protein zero (MPZ). Genomics 1993 Sep;17(3):755-8).

Epilepsy affects approximately 1% of the population making it one of the most common neurological diseases. There are many forms of epilepsy - most are rare. To date, twelve forms of epilepsy have been demonstrated to possess some genetic basis. For example, LaFora Disease (progressive myclonic, type 2) is thought to result from a mutation in the EPM2A gene, which is located on chromosome 6. This gene is thought to produce laforin, a protein similar to a group of protein-tyrosine phosphatases that help maintain a balance of sugars in the blood stream. Too much laforin destroys brain cells, which may then lead to the development of LaFora Disease. Much progress has been made in narrowing down regions of chromosomes associated with different forms of epilepsy. This invention will be particularly useful in helping to further the research and diagnostics of this disease (Minassian BA, Lee JR, Herbrick JA, Huizenga J, Soder S, Mungall AJ, Dunham I, Gardner R, Fong CY, Carpenter S, Jardim L, Satishchandra P, Andermann E, Snead OC 3rd, Lopes-Cendes I, Tsui LC, Delgado-Escueta AV, Rouleau GA, Scherer SW. Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nat Genet 1998 Oct;20(2):171-4).

Tremor, or uncontrollable shaking, is a common symptom of neurological disorders such as Parkinson disease, head trauma and stroke. However, many people with tremor have what is called idiopathic or essential tremor. In these cases, which number 3-4 million people in the US, the tremor itself is the only symptom of the disorder. In more than half of cases, essential tremor is inherited as an autosomal dominant trait, which means that children of an affected individual will have a 50 percent chance of also developing the disorder. In 1997, the ETM1 gene (also called FET1) was mapped to chromosome 3 in a study of Icelandic families, while another gene, called ETM2, was mapped to chromosome 2. That two genes for essential tremor have been found on two different chromosomes demonstrates that mutations in a variety of genes lead to essential tremor (Gulcher JR, Jonsson P, Kong A, Kristjansson K, Frigge ML, Karason A, Einarsdottir IE, Stefansson H, Einarsdottir AS, Sigurthoardottir S, Baldursson S, Bjomsdottir S, Hrafnkelsdottir SM, Jakobsson F, Benedickz J, Stefansson K. Mapping of a familial essential tremor gene, FET1, to chromosome 3q13. Nat Genet 1997 Sep;17(1):84-7).

Fragile X syndrome is the most common inherited form of mental retardation currently known. Fragile X syndrome is a defect in the X chromosome and its effects are seen more frequently, and with greater severity, in males than females. In normal individuals, the FMR1 gene is transmitted stably from parent to child. However, in Fragile X individuals, there is a mutation in one end of the gene (the 5' untranslated region), consisting of an amplification of a CGG repeat. Patients with fragile X syndrome have 200 or more copies of the CGG motif. The huge expansion of this repeat means that the FMR1 gene is not expressed, so no FMR1 protein is made. Although the exact function of FMR1 protein in the cell is unclear, it is known that it binds RNA. A similar nucleotide repeat expansion is seen in other diseases, such as Huntington disease (Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. Cell 1993 Jul 30;74(2):291-8). The remaining 5% of fragile X cases correspond to other molecular alterations in FMR1 gene such as partial or complete deletions, or point mutations within the coding sequence (Castellvi-Bel S, Sanchez A, Badenas C, Mallolas J, Barcelo A, Jimenez D, Villa M, Estivill X, Mila M. Single-strand conformation polymorphism analysis in the FMR1 gene. Am J Med Genet 1999 May 28;84(3):262-5).

Friedreich's ataxia (FRDA) is a rare inherited disease characterized by the progressive loss of voluntary muscular coordination (ataxia) and heart enlargement. FRDA is an autosomal recessive disease caused by a mutation of a gene called frataxin, which is located on chromosome 9. This mutation means that there are many extra copies of a DNA segment, the trinucleotide GAA. A normal individual has 8 to 30 copies of this trinucleotide, while FRDA patients have as many as 1,000 (Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 1996 Mar 8;271(5254):1423-7).

Huntington disease (HD) is an inherited, degenerative neurological disease that leads to dementia. About 30,000 Americans have HD and about 150,000 more are at risk of inheriting the disease from a parent. The HD gene, whose mutation results in Huntington disease, was mapped to chromosome 4 in 1983 and cloned in 1993. The mutation is a characteristic expansion of a nucleotide triplet repeat in the DNA that codes for the protein huntingtin. The number of repeated triplets - CAG (cytosine, adenine, guanine) - increases with the age of the patient. Since people who have those repeats always suffer from Huntington disease, it suggests that the mutation causes a gain-of-function, in which the mRNA or protein takes on a new property or is expressed inappropriately. With the discovery of the HD gene, a new predictive test was developed that allows those at risk to find out whether or not they will develop the disease (The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993 Mar 26;72(6):971-83).

There are three separate diseases that carry the name Niemann-Pick: Type A is the acute infantile form, Type B is a less common, chronic, non-neurological form, while Type C is a biochemically and genetically distinct form of the disease. Recently, the major locus responsible for Niemann-Pick type C (NP-C) was cloned from chromosome 18, and found to be similar to proteins that play a role in cholesterol homeostasis. Usually, cellular cholesterol is imported into lysosomes. Cells taken from NP-C patients have been shown to be defective in releasing cholesterol from lysosomes. This leads to an excessive build-up of cholesterol inside lysosomes, causing processing errors. NPC1 was found to have known sterol-sensing regions similar to those in other proteins, which suggests it plays a role in regulating cholesterol traffic (Carstea ED, Morris JA, Coleman KG, Loftus SK, Zhang D, Cummings C, Gu J, Rosenfeld MA, Pavan WJ, Krizman DB, Nagle J, Polymeropoulos MH, Sturley SL, Ioannou YA, Higgins ME, Comly M, Cooney A, Brown A, Kaneski CR, Blanchette-Mackie EJ, Dwyer NK, Neufeld EB, Chang TY, Liscum L, Tagle DA, et al. Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. Science 1997 Jul 11;277(5323):228-31).

Parkinson disease is a neuodegenerative disease that manifests as a tremor, muscular stiffness and difficulty with balance and walking. A candidate gene for some cases of Parkinson disease was mapped to chromosome 4. Mutations in this gene have now been linked to several Parkinson disease families. The product of this gene, a protein called alpha-synuclein, is a familiar culprit: a fragment of it is a known constituent of Alzheimer disease plaques. Since alpha-synuclein fragments are implicated in both Parkinson and Alzheimer diseases (Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997 Jun 27;276(5321):2045-7).

Spinocerebellar atrophy, of which there are several types, is a degeneration of the spinal cord and the cerebellum, the small fissured mass at the base of the brain, behind the brain stem. The cerebellum is concerned with coordination of movements, so atrophy or "wasting away" of this critical control center results in a loss of muscle coordination. The basic defect in all types of spinocerebellar atrophy is a an expansion of a CAG triplet repeat. In this way, it is similar to fragile-X syndrome, Huntington disease and myotonic dystrophy, all of which exhibit a triplet repeat expansion of a gene. In the case of spinocerebellar atrophy I, the gene is SCA1, found on chromosome 6 (Banfi S, Servadio A, Chung MY, Kwiatkowski TJ Jr, McCall AE, Duvick LA, Shen Y, Roth EJ, Orr HT, Zoghbi HY. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat Genet 1994 Aug;7(4):513-20).

Williams syndrome is a rare congenital disorder characterized by physical and development problems. In Williams syndrome individuals, both the gene for elastin and an enzyme called LIM kinase are deleted. Both genes map to the same small area on chromosome 7. In normal cells, elastin is a key component of connective tissue, conferring its elastic properties. Mutation or deletion of elastin lead to the vascular disease observed in Williams syndrome. On the other hand, LIM kinase is strongly expressed in the brain, and deletion of LIM kinase is thought to account for the impaired visuospatial constructive cognition in Williams syndrome. Williams syndrome is a contigious disease, meaning that the deletion of this section of chromosome 7 involves several more genes (Lenhoff HM, Wang PP, Greenberg F, Bellugi U. Williams syndrome and the brain. Sci Am 1997 Dec;277(6):68-73).

6.5. Applications in Various Metabolic Diseases

Adrenoleukodystrophy (ALD) is a rare, inherited metabolic disorder wherein myelin sheath on nerve fibers in the brain is lost, and the adrenal gland degenerates, leading to progressive neurological disability and death. ALD gene was discovered in 1993 and corresponding protein is ember of a family of transporter proteins (Mosser J, Douar AM, Sarde CO, Kioschis P, Feil R, Moser H, Poustka AM, Mandel JL, Aubourg P. Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. Nature 1993 Feb 25;361(6414):726-30).

Atherosclerosis is a disease that can affect people at any age, although it usually doesn't pose a threat until people reach their forties or fifties. It is characterized by a narrowing of the arteries caused by cholesterol-rich plaques of immune-system cells. Key risk factors for atherosclerosis, which can be genetic and/or environmental. A protein called apolipoprotein E, which can exist in several different forms, is coded for by a gene found on chromosome 19. Detection of defects in apolipoprotein E gene is useful for diagnosis and treatment atherosclerosis (Wenham PR, Newton CR, Price WH. Analysis of apolipoprotein E genotypes by the Amplification Refractory Mutation System. Clin Chem 1991 Feb;37(2):241-4).

Gaucher disease is characterized by the accumulation of glucocerebroside, leading to enlargement of the liver and spleen and lesions in the bones. It is caused by an inherited deficiency of the enzyme glucocerebrosidase. Many mutations exist, but four of these account for over 97% of the mutations in Ashkenazi Jews, the population group in which Gaucher disease is the most common. Detection of genetic mutations is useful for genetic counseling and gene therapy (Beutler E. Gaucher disease: new molecular approaches to diagnosis and treatment. Science 1992 May 8;256(5058):794-9).

Glanzmann's thrombasthenia (GT) arises from a qualitative or quantitative defect in the GPIIb-IIIa complex (integrin alphaIIbeta3a), the mediator of platelet aggregation (Ruan J, Schmugge M, Clemetson KJ, Cazes E, Combrie R, Bourre F, Nurden AT. Homozygous Cys542-->Arg substitution in GPIIIa in a swiss patient with type I Glanzmann's thrombasthenia. Br J Haematol 1999 May;105(2):523-31).

Gyrate atrophy of the choroid and retina leads to a progressive loss of vision, with total blindness usually occurring between the ages of 40 and 60. The disease is an inborn error of metabolism. The gene whose mutation causes gyrate atrophy is found on chromosome 10, and encodes ornithine ketoacid aminotransferase (OAT) enzyme. Different inherited mutations in OAT cause differences in the severity of symptoms of the disease (Akaki Y, Hotta Y, Mashima Y, Murakami A, Kennaway NG, Weleber RG, Inana G. A deletion in the ornithine aminotransferase gene in gyrate atrophy. J Biol Chem 1992 Jun 25;267(18):12950-4; O'Donnell JJ, Vannas-Sulonen K, Shows TB, Cox

DR. Gyrate atrophy of the choroid and retina: assignment of the ornithine aminotransferase structural gene to human chromosome 10 and mouse chromosome 7. Am J Hum Genet 1988 Dec;43(6):922-8).

Diabetes is a chronic metabolic disorder that adversely affects the body's ability to manufacture and use insulin, a hormone necessary for the conversion of food into energy. The disease greatly increases the risk of blindness, heart disease, kidney failure, neurological disease and other conditions for the pproximately 16 million Americans who are affected by it. Type I, or juvenile onset diabetes, is the more severe form of the illness. Type I diabetes is what is known as a 'complex trait', which means that mutations in several genes likely contribute to the disease. About 10 loci in the human genome have now been found that seem to confer susceptibility to Type I diabetes. Among these are (1) a gene at the locus IDDM2 on chromosome 11 and (2) the gene for glucokinase (GCK), an enzyme that is key to glucose metabolism which helps modulate insulin secretion, on chromosome 7 (Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, et al. A genome-wide search for human type 1 diabetes susceptibility genes. Nature 1994 Sep 8;371(6493):130-6).

Genetic mutations in the coding or exon-intron regions of the uncoupling protein 2 (UCP2) gene have been associated with non-insulin-dependent diabetes mellitus (NIDDM). (Shiinoki T, Suehiro T, Ikeda Y, Inoue M, Nakamura T, Kumon Y, Nakauchi Y, Hashimoto K. Screening for variants of the uncoupling protein 2 gene in Japanese patients with non-insulin-dependent diabetes mellitus. Metabolism 1999 May;48(5):581-4).

NeuroD/BETA2 gene mapped to the long arm of human chromosome 2 (2q32) where the IDDM7 gene has previously been mapped, involved in type 1 but not type 2 diabetes (Iwata I, Nagafuchi S, Nakashima H, Kondo S, Koga T, Yokogawa Y, Akashi T, Shibuya T, Umeno Y, Okeda T, Shibata S, Kono S, Yasunami M, Ohkubo H, Niho Y. Association of polymorphism in the NeuroD/BETA2 gene with type 1 diabetes in the Japanese. Diabetes 1999 Feb;48(2):416-9).

The mutation in Shc adaptor proteins's gene useful as a marker for identifying impaired insulin secretion, insulin resistance, and type 2 diabetes mellitus (Almind K, Ahlgren MG, Hansen T, Urhammer SA, Clausen JO, Pedersen O. Discovery of a Met300Val variant in Shc and studies of its relationship to birth weight and length, impaired insulin secretion, insulin resistance, and type 2 diabetes mellitus. J Clin Endocrinol Metab 1999 Jun;84(6):2241-4).

Obesity is an excess of body fat that frequently results in a significant impairment of health. Doctors generally agree that men with more than 25% body fat and women with more than 30% are obese. Obesity is a known risk factor for chronic diseases including heart disease, diabetes, high blood pressure, stroke and some forms of cancer. Evidence suggests that obesity has more than one cause: genetic, environmental, psychological and other factors may all play a part. The hormone leptin, produced by adipocytes appears to be the key factor and its Ob gene was mapped to chromosome 7. A whole network of signals contributes to weight homeostasis, and other key players are being discovered on an ongoing basis. Since the market for effective weight-reducing therapies is enormous, this gene has important implications for predicting and treating obesity (Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. Nature 1994 Dec 1;372(6505):425-32).

Paroxysmal nocturnal hemoglobinuria (PNH) is associated with a high risk of major thrombotic events, most commonly thrombosis of large intra-abdominal veins. Most patients who die of their disease die of thrombosis. PNH blood cells are deficient in an enzyme known as PIG-A, which is required for the biosynthesis of cellular anchors. The PIG-A gene is found on the X chromosome. Although not an inherited disease, PNH is a genetic disorder, known as an acquired or somatic genetic disorder (Bessler M, Mason PJ, Hillmen P, Miyata T, Yamada N, Takeda J, Luzzatto L, Kinoshita T. Paroxysmal nocturnal haemoglobinuria (PNH) is caused by somatic mutations in the PIG-A gene. EMBO J 1994 Jan 1;13(1):110-7).

Phenylketonuria (PKU) is an inherited error of metabolism caused by a deficiency in the enzyme phenylalanine hydroxylase. Loss of this enzyme results in mental retardation, organ damage, unusual posture and can, in cases of maternal PKU, severely compromise pregnancy. Classical PKU is an autosomal recessive disorder, caused by mutations in both alleles of the gene for phenylalanine hydroxylase (PAH), found on chromosome 12. In some cases, mutations in PAH will result in a phenotypically mild form of PKU called hyperphenylalanemia. Both diseases are the result of a variety of mutations in the PAH locus (DiLella AG, Marvit J, Brayton K, Woo SL. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. Nature 1987 May 28-Jun 3;327(6120):333-6).

Refsum disease is a rare disorder of lipid metabolism that is inherited as a recessive trait. Symptoms may include a degenerative nerve disease (peripheral neuropathy), failure of muscle coordination (ataxia), retinitis pigmentosa (a progressive vision disorder), and bone and skin changes. Refsum disease is characterized by an accumulation of phytanic acid in the plasma and tissues. In 1997 the gene for Refsum disease was identified and mapped to chromosome 10. The protein product of the gene, PAHX, is an enzyme that is required for the metabolism of phytanic acid (Jansen GA, Ofman R, Ferdinandusse S, Ijlst L, Muijsers AO, Skjeldal OH, Stokke O, Jakobs C, Besley GT, Wraith JE, Wanders RJ. Refsum disease is caused by mutations in the phytanoyl-CoA hydroxylase gene. Nat Genet 1997 Oct;17(2):190-3).

X-linked liver glycogenosis (XLG) is probably the most frequent glycogen-storage disease. XLG can be divided into two subtypes: XLG I, with a deficiency in phosphorylase kinase (PHK) activity in peripheral blood cells and liver; and XLG II, with normal in vitro PHK activity in peripheral blood cells and with variable activity in liver. Both types of XLG are caused by mutations in the same gene, PHKA2, that encodes the regulatory alpha subunit of PHK (Hendrickx J, Lee P, Keating JP, Carton D, Sardharwalla IB, Tuchman M, Baussan C, Willems PJ. Complete genomic structure and mutational spectrum of PHKA2 in patients with x-linked liver glycogenosis type I and II. Am J Hum Genet 1999 Jun;64(6):1541-9).

Cystic fibrosis (CF) is the most common fatal genetic disease in the US today. It causes the body to produce a thick, sticky mucus that clogs the lungs, leading to infection, and blocks the pancreas, stopping digestive enzymes from reaching the intestines where they are required to digest food. CF is caused by a defective gene, which codes for a sodium and chloride transporter found on the surface of the epithelial cells. Several hundred mutations have been found in this gene, all of which result in defective transport of salt ions. CF research has accelerated sharply since the discovery of CFTR in 1989. Gene therapy was then tried on a limited number of CF patients (Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989 Sep 8;245(4922):1066-73).

Diastrophic dysplasia (DTD) is a rare growth disorder in which patients are usually short, have club feet and have malformed hands and joints. The gene whose mutation results in DTD maps to chromosome 5 and encodes a novel sulfate transporter. This ties in with the observation of unusual concentrations of sulfate in various tissues of DTD patients (Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 1994 Sep 23;78(6):1073-87).

Long-QT syndrome (LQTS) results from structural abnormalities in the potassium channels of the heart, which predispose affected persons to an accelerated heart rhythm (arrhythmia). This can lead to sudden loss of consciousness and cause sudden cardiac death in teenagers and young adults who are faced with stressors ranging from exercise to loud sounds. LQTS is usually inherited as an autosomal dominant trait. Gene LQT1 has been mapped to chromosome 11 and mutations lead to serious structural defects in the person's cardiac potassium channels. There also appear to be other genes, tentatively located on chromosomes 3, 6 and 11 whose mutated products contribute to, or cause, LQT syndrome (Barhanin J, Lesage F, Guillemare E, Fink M, Lazdunski M, Romey G. K(V)LQT1 and IsK (minK) proteins associate to form the I(Ks) cardiac potassium current. Nature 1996 Nov 7;384(6604):78-80).

Menkes' syndrome is an inborn error of metabolism that markedly decreases the cells' ability to absorb copper. The disorder causes severe cerebral degeneration and arterial changes, resulting in death in infancy. Menkes' disease is transmitted as an X-linked recessive trait. A number of other diseases, including type IX Ehlers-Danlos syndrome, are the result of allelic mutations (i.e. mutations in the same gene, but having slightly different symptoms) and research into these genes is useful in finding cure (Chelly J, Tumer Z, Tonnesen T, Petterson A, Ishikawa-Brush Y, Tommerup N, Horn N, Monaco AP. Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. Nat Genet 1993 Jan;3(1):14-9).

Pendred syndrome (osteochondrodysplasia) is an inherited disorder that accounts for as much as 10% of hereditary deafness. Patients usually also suffer from thyroid goiter. The normal gene makes a protein, called pendrin, that is found at significant levels only in the thyroid and is closely related to a number of sulfate transporters. When the gene for this protein is mutated, the person carrying it will exhibit the symptoms of Pendred syndrome. The discovery of pendrin should also stimulate new angles of research into thyroid physiology and the role of altered sulfur transport in human disease (Kopp P. Pendred's syndrome: identification of the genetic defect a century after its recognition. Thyroid 1999 Jan;9(1):65-9).

Adult polycystic kidney disease (APKD) is characterized by large cysts in one or both kidneys and a gradual loss of normal kidney tissue. Patients with APKD can die from renal failure, or from the consequences of hypertension (high arterial blood pressure). In 1994 the European Polycystic Kidney Disease Consortium isolated a gene from chromosome 16 that was disrupted in a family with APCD. The protein encoded by the PKD1 gene is an integral membrane protein involved in cell-cell interactions and cell-matrix interactions. The role of PKD1 in the normal cell linked to microtubule-mediated functions, such as the placement of Na(+), K(+)-ATPase ion pumps in the membrane. Programmed cell death, or apoptosis, is also invoked in APKD (The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. Cell 1995 Apr 21;81(2):289-98).

Wilson's disease is a rare autosomal recessive disorder of copper transport, resulting in copper accumulation and toxicity to the liver and brain. The cornea of the eye can also be affected: the 'Kayser-Fleischer ring' is a deep copper-colored ring at the periphery of the cornea. The gene for Wilson's disease (ATP7B) was mapped to chromosome 13. The sequence of the gene was found to be similar to sections of the gene defective in Menkes disease. These genes will be useful for studying copper transport and liver pathophysiology, and are useful in the development of a therapy for Wilson disease (Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW. The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nat Genet 1993 Dec;5(4):327-37).

Zellweger syndrome (ZS) is a hereditary disorder affecting infants, and usually results in death. Unusual problems in prenatal development, an enlarged liver, high levels of iron and copper in the blood, and vision disturbances are among the major manifestations. The PXR1 gene has been mapped to chromosome 12; mutations in this gene cause ZS. The PXR1 gene product is a receptor found on the surface of peroxisomes - microbodies that carry out a number of metabolically important reactions such as cellular lipid metabolism and metabolic oxidations (Marynen P, Fransen M, Raeymaekers P, Mannaerts GP, Van Veldhoven PP. The gene for the peroxisomal targeting signal import receptor

(PXR1) is located on human chromosome 12p13, flanked by TPI1 and D12S1089. Genomics 1995 Nov 20;30(2):366-8). The gene mutation of peroxisome proliferator-activated receptor gamma (PPARgamma) protein associated with bone mineral density (BMD) and osteoporosis in postmenopausal women (Ogawa S, Urano T, Hosoi T, Miyao M, Hoshino S, Fujita M, Shiraki M, Orimo H, Ouchi Y, Inoue S. Association of Bone Mineral Density with a Polymorphism of the Peroxisome Proliferator-Activated Receptor gamma Gene: PPARgamma Expression in Osteoblasts. Biochem Biophys Res Commun 1999 Jun 24;260(1):122-126).

Paramyotonia congenita is a temperature-sensitive skeletal muscle disorder caused by missense mutations that occur in the adult skeletal muscle voltage-gated sodium channel genes (Bendahhou S, Cummins TR, Kwiecinski H, Waxman SG, Ptacek LJ. Characterization of a new sodium channel mutation at arginine 1448 associated with moderate paramyotonia congenita in humans. Nat Genet 1999 Jun;22(2):164-7).

6.6. Application in Disorders Associated with Cellular Signaling

Ataxia telangiectasia (A-T) is a progressive, degenerative disease characterized by cerebellar degeneration, immunodeficiency, radiosensitivity (sensitivity to radiant energy, such as x-ray) and a predisposition to cancer. Back in 1988 the gene responsible for A-T was mapped to chromosome 11. The diverse symptoms seen in A-T reflect the main role of ATM, which is to induce several cellular responses to DNA damage. When the ATM gene is mutated, these signaling networks are impaired and so the cell does not respond correctly to minimize the damage (Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 1995 Jun 23;268(5218):1749-53). Mutations resulting in defective splicing in patients with AT can be detected by the protein-truncation assay followed by sequence analysis of genomic DNA. These splicing mutations led to a variety of genetic effects, including exon skipping and intron retention, activation of cryptic splice sites, or creation of new splice sites. Such mutations are detected according to principles disclosed by Teraoka et al., (Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA, Concannon P. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet 1999 Jun;64(6):1617-31). More mutations associated with A-T are found in U. S. Pat. No. 5,858,661 issued to Shiloh on January 12, 1999 and incorporated herein by way of reference.

Baldness. 5-alpha reductase is an enzyme that was first discovered in the prostate. Here, it catalyzes the conversion of testosterone to dihydrotestosterone, which in turn binds to the androgen receptor and initiates development of the external genitalia and prostate. The gene for 5-alpha reductase has been mapped to chromosome 5. More recently, 5-alpha reductase was found in human scalp and elsewhere in the skin, where it carries out the same reaction as in the prostate. It is thought that disturbances in 5-alpha reductase activity in skin cells contribute to male pattern baldness, acne or hirsutism (Jenkins EP, Hsieh CL, Milatovich A, Normington K, Berman DM, Francke U, Russell DW. Characterization and chromosomal mapping of a human steroid 5 alpha-reductase gene and pseudogene and mapping of the mouse homologue. Genomics 1991 Dec;11(4):1102-12).

Cockayne syndrome is a rare inherited disorder in which people are sensitive to sunlight, have short stature and have the appearance of premature aging. In the classical form of Cockayne syndrome (Type I), the symptoms are progressive and typically become apparent after the age of one year. An early onset or congenital form of Cockayne syndrome (Type II) is apparent at birth. Unlike other DNA repair diseases, Cockayne syndrome is not linked to cancer. Two genes defective in Cockayne syndrome, CSA and CSB, have been identified so far. Both genes code for proteins that interacts with components of the transcriptional machinery and with DNA repair proteins. Defects in the XPB, XPD, and XPG genes can result in three different syndromes Cockayne syndrome, xeroderma pigmentosum, or trichothiodystrophy, depending on the specific mutation involved (van Gool AJ, van der Horst GT, Citterio E, Hoeijmakers JH. Cockayne syndrome: defective repair of transcription? EMBO J 1997 Jul 16; 16 (14):4155-62).

Glaucoma is a general term used for a group of diseases that can lead to damage to the eye's optic nerve and result in blindness. The most common form of the disease is open-angle glaucoma, which affects about three million Americans, half of whom don't know they have it. Glaucoma has no symptoms at first but over the years can steal its victims' sight, with side vision being effected first. It is estimated that nearly 100,000 individuals in the US suffer from glaucoma due to a mutation in the GLC1A gene, found on chromosome 1. The development of tests for the early detection of the disease, as well as providing a basis for research into effective therapies will depend on detection of mutations in GLC1A and other genes involved in subtypes of glaucoma (Stone EM, Fingert JH, Alward WLM, Nguyen TD, Polansky JR, Sunden SLF, Nishimura D, Clark AF, Nystuen A, Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC. Identification of a gene that causes primary open angle glaucoma. Science 1997 Jan 31;275(5300):668-70). Primary congenital glaucoma (PCG) is an autosomal recessive eye disease that associates with GLC3A locus on 2p21. At this locus, mutations in the cytochrome P4501B1 (CYP1B1) gene are identified as a molecular basis for this condition (Plasilova M, Stoilov I, Sarfarazi M, Kadasi L, Ferakova E, Ferak V. Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma. J Med Genet 1999 Apr;36(4):290-4).

Two forms of autosomal-dominant lattice corneal dystrophy (LCD), types I and IIIA, and late-onset LCD have been shown to be caused by different mutations within the transforming growth factor, beta-induced (TGFBI) gene. Among several mutations sequence changes within exon 14 of the TGFBI gene on chromosome 5q31, at codon 622, and at codon 626, are presumed to be responsible for the disease (Stewart H, Black GC, Donnai D, Bonshek RE, McCarthy J, Morgan S, Dixon MJ, Ridgway

AA. A mutation within exon 14 of the TGFBI (BIGH3) gene on chromosome 5q31 causes an asymmetric, late-onset form of lattice corneal dystrophy. Ophthalmology 1999 May;106(5):964-70).

SRY (sex-determining region Y gene) is found of the Y chromosome. In the cell, it binds to DNA and in doing so distorts it dramatically out of shape. This alter the properties of the DNA and likely alters the expression of a number of genes, leading to testis formation. Therefore XX men who lack a Y chromosome also lack SRY and frequently do not develop secondary sexual characteristics in the usual way. This has been particularly important in discovering the interactions of SRY with other genes in sex determination (Berta P, Hawkins JR, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. Genetic evidence equating SRY and the testis-determining factor. Nature 1990 Nov 29;348(6300):448-50).

Tuberous sclerosis is an hereditary disorder characterized by benign, tumor-like nodules of the brain and/or retinas, skin lesions, seizures and/or mental retardation. Patients may experience a few or all of the symptoms with varying degrees of severity. Two loci for tuberous sclerosis have been found: TSC1 on chromosome 9, and TSC2 on chromosome 16. It took four years to pin down a specific gene from the TSC1 region of chromosome 9: in 1997, a promising candidate was found. Called hamartin by the discoverers, it is similar to a yeast protein of unknown function, and appears to act as a tumor suppressor: without TSC1, growth of cells proceeds in an unregulated fashion, resulting in tumor formation. TSC2 codes for a protein called tuberin, which, through database searches, was found to have a region of homology to a protein (GAP3, a GTPase-activation protein) found in pathways that regulate the cell (The European Chromosome 16 Tuberous Sclerosis Consortium. Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 1993 Dec 31;75(7):1305-15; van Slegtenhorst M, de Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, Lindhout D, van den Ouweland A, Halley D, Young J, Burley M, Jeremiah S, Woodward K, Nahmias J, Fox M, Ekong R, Osborne J, Wolfe J, Povey S, Snell RG, Cheadle JP, Jones AC, Tachataki M, Ravine D, Kwiatkowski DJ, et al. Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. Science 1997 Aug 8;277(5327):805-8).

Deletions involving chromosome 9 occur in more than 50% of human bladder cancers of all grades and stages. A critical region of deletion at 9q34 between the markers D9S149 and D9S66, an interval which contains the TSC1gene acts as a bladder tumour suppressor gene (Hornigold N, Devlin J, Davies AM, Aveyard JS, Habuchi T, Knowles MA. Mutation of the 9q34 gene TSC1 in sporadic bladder cancer. Oncogene 1999 Apr 22;18(16):2657-61). Testing other bladder cancer-associated gene like E2F-1 is also useful (Rabbani F, Richon VM, Orlow I, Lu ML, Drobnjak M, Dudas M, Charytonowicz E, Dalbagni G, Cordon-Cardo C. Prognostic significance of transcription factor E2F-1 in bladder cancer: genotypic and phenotypic characterization. J Natl Cancer Inst 1999 May 19;91(10):874-81).

Waardenburg syndrome (WS) includes a wide bridge of the nose, pigmentary disturbances such as two different colored eyes, white forelock and eyelashes and premature graying of the hair, and some degree of cochlear deafness. The several types of WS are inherited in dominant fashion. The human gene on chromosome 2 is the same as mouse Pax3, which is one of a family of eight mouse Pax genes that are involved in regulating embryonic development at the level of transcription (Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P, Strachan T. Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. Nature 1992 Feb 13;355(6361):635-6).

Werner syndrome (WS) is a premature aging disease. Its physical characteristics may include short stature (common from childhood on) and other features usually developing during adulthood: wrinkled skin, baldness, cataracts, muscular atrophy and a tendency to diabetes mellitus, among others. The disorder is inherited and transmitted as an autosomal recessive trait. The gene for Werner disease (WRN) was mapped to chromosome 8 and cloned: by comparing its sequence to existing sequences in GenBank, it is a predicted helicase (DNA unwinder important for DNA replication) belonging to the RecQ family. The recent cloning of the genes involved in Bloom syndrome (BLM) and Werner syndrome (WRN) show that both are DNA and RNA helicases with homology to each other and to other DExH box helicases. Other phenotypically distinctive disorders caused by different helicase mutations include Cockayne syndrome, xeroderma pigmentosum and trichothiodystrophy. Diseases such as progeria and Mulvihill-Smith syndrome resemble WS. The molecular role of WRN in Werner syndrome is thought to be important in the aging process in general (Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL, Martin GM, Oshima J, Loeb LA. The Werner syndrome protein is a DNA helicase. Nat Genet 1997 Sep;17(1):100-3).

Miscellaneous other genes involved in clinical disorders are also recognized as useful and include but are not limited to beta-globin, phenylalanine hydroxylase, alpha-antitrypsin, 21-hydroxylase, pyruvate dehydrogenase, dihydropteridine reductase, rhodopsin, nerve growth factor, superoxide dismutase, adenosine deaminase, beta-thalassemia, ornithine transcarbamylase, collagen, beta-hexosaminidase, topoisomerase II, hypoxanthine phosphoribosyltransferase, phenylalanine 4-monooxygenase, Factor VIII, Factor IX, nucleoside phosphorylase, glucose-6-phosphate dehydrogenase, phosphoribosyltransferase.

7.7. Applications Relating to Infectious and Vector-Borne Diseases

SNP mutation in transcription factor OCT-1 and associated with fourfold increased susceptibility to cerebral malaria (Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, Marsh K, Kwiatkowski D. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nat Genet 1999 Jun;22(2):145-50).

Creutzfeldt-Jakob disease (CJD) belongs to a group of prion diseases that may be infectious, sporadic, or hereditary. The 200K point mutation in the PRNP gene is the most frequent cause of hereditary CJD, accounting for >70% of families with CJD worldwide. A major haplotype is involved in Sephardic migrants expelled from Spain in the Middle Ages. Slovakian families and a family of Polish origin show another unique haplotype. Thus, the founder effect and independent mutational events are responsible for the current geographic distribution of hereditary CJD associated with the 200K mutation (Lee HS, Sambuughin N, Cervenakova L, Chapman J, Pocchiari M, Litvak S, Qi HY, Budka H, del Ser T, Furukawa H, Brown P, Gajdusek DC, Long JC, Korczyn AD, Goldfarb LG. Ancestral Origins and Worldwide Distribution of the PRNP 200K Mutation Causing Familial Creutzfeldt-Jakob Disease. Am J Hum Genet 1999 Apr;64(4):1063-1070). Several neurodegenerative diseases, including kuru, scrapie, and bovine spongiform encephalopathy (BSE) are also thought to involve similar mutations.

Gene-based approach is useful for detecting drug-resistance and identifying better therapeutic modalities against infectious diseases like tuberculosis, malaria, or HIV (Cooksey RC, Morlock GP, McQueen A, Glickman SE, Crawford JT. Characterization of streptomycin resistance mechanisms among Mycobacterium tuberculosis isolates from patients in New York City. Antimicrob Agents Chemother 1996; 40(5):1186-8; Wellems TE, Walker-Jonah A, Panton LJ. Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7. Proc Natl Acad Sci U S A 1991 Apr 15;88(8):3382-6). U. S. Pat. 5,861,242 to Chee, et al., on January 19, 1999 discloses an array of nucleic acid probes on biological chips for diagnosis of human immunodeficiency virus (HIV). Chambers et al., (Chambers J, Angulo A, Amaratunga D, Guo H, Jiang Y, Wan JS, Bittner A, Frueh K, Jackson MR, Peterson PA, Erlander MG, Ghazal P. DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. J Virol 1999 Jul;73(7):5757-66) describe viral DNA chip for simultaneous expression measurements of nearly all known open reading frames (ORFs) in the largest member of the herpesvirus family, human cytomegalovirus (HCMV). In this study, an HCMV chip was fabricated and used to characterize the temporal class of viral gene expression.

8.8. Identification of Genetic Distance

Microarray-based assays allows rapid comparative sequence analysis of intra- and interspecies genetic variation, i.e., genomic epidemiology studies (Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. This type of application is useful for forensic and legal purposes (e.g., pedigree search) as a method of DNA fingerprinting.

8.9 Other Refer nces

The references disclosed infra deal with additional technical and general aspects of the present invention and are expressly incorporated in their entirety in this disclosure by way of reference.

Baxter, G. et al., "Microfabrication in Silicon Microphysiometry, " Clin. Chem., vol. 40, No. 9, 1800-1804 (1994).

Beattie, K. et al., "Advances in Genosensor Research, " Clin. Chem., vol. 41, No. 5, 700-706 (1995).

Brenner, S. and Livak, K., "DNA Fingerprinting by Sampled Sequencing, " Proc. Natl. Acad. Sci. USA, vol. 86, 8902-8906 (1989).

Broude, N. et al., "Enhanced DNA Sequencing by Hybridization, " Proc. Natl. Acad. Sci. USA, vol. 91, 3072-3076 (1994).

Burns, M. et al., "Microfabricated Structures for Integrated DNA Analysis, " Proc. Natl. Acad. Sci. USA, vol. 93, 5556-5561 (1996).

Caetano-Anolles, G. et al., "Primer-Template Interactions During DNA Amplification Fingerprinting with Single Arbitrary Oligonucleotides, " Mol. Gen. Genet., vol. 235, 157-165 (1992).

Canard, B. and Sarfati, R.S., "DNA Polymerase Fluorescent Substrates with Reversible 3'-tags, " Gene, vol. 148, 1-6 (1994).

Carrano, A.V. et al., "A High-Resolution, Fluorescence-Based, Semiautomated Method for DNA Fingerprinting, " Genomics, vol. 4, 129-136 (1989).

Cheng, J. et al., "Chip PCR. II. Investigation of Different PCR Amplification Systems in Microfabricated Silicon-Glass Chips, " Nucleic Acids Research, vol. 24, No. 2, 380-385 (1996).

Chetverin, A. and Kramer, F., "Oligonucleotide Arrays: New Concepts and Possibilities, " BioTechnology, vol. 12, 1093-1099 (1994).

Davis, L. et al., "Rapid DNA Sequencing Based Upon Single Molecule Detection, " Genetic Analysis, Techniques, and Applications, vol. 8, No. 1, 1-7 (1991).

Drmanac, R. et al., "DNA Sequence Determination by Hybridization: A Strategy for Effecient Large-Scale Sequencing, " Science, vol. 260, 1649-1652 (1993).

Eggers, M. and Ehrlich, D., "A Review of Microfabricated Devices for Gene-Based Diagnostics, " Hematologic Pathology, vol. 9, No. 1, 1-15 (1995).

Eggers, M. et al., "A Microchip for Quantitative Detection of Molecules Utilizing Luminescent and Radioisotope Reporter Groups, " BioTechniques, vol. 17, No. 3, 516-524 (1994).

Gibbs, R. et al., "Identification of Mutations Leading to the Lesch-Nyhan Syndrome Automated Direct DNA Sequencing of In Vitro Amplified cDNA, " Proc. Natl. Acad. Sci. USA, vol. 86, 1919-1923 (1989).

Green, E. and Green, P., "Sequence-tagged Site (STS) Content Mapping of Human Chromosomes: Theoretical Considerations and Early Experiences, " PCR Methods and Applications, vol. 1, 77-90 (1991).

Gyllensten, U. and Allen, M., "PCR-based HLA Class II Typing, " PCR Methods and Applications, vol. 1, 91-98 (1991).

Gyllensten, U. and Erlich, H., "Generation of Single-Stranded DNA by the Polymerase Chain Reaction and its Application to Direct Sequencing of the HLA-DQA Locus, " Proc. Natl. Acad. Sci. USA, vol. 85, 7652-7656 (1988).

Han, J. and Rutter, W., ".lambda.gt22S, a Phage Expression Vector for the Directional Cloning of cDNA by the use of a Single Restriction Enzyme SfiI, " Nucleic Acids Research, vol. 16, No. 24, 11837 (1988).

Jacobs, K. et al., "The Thermal Stability of Oligonucleotide Duplexes is Sequence Independent in Tetraalkylammonium Salt Solutions: Application to Identifying Recombinant DNA Clones, " Nucleic Acids Res., vol. 16, 4637-4650 (1988).

Kikuchi, Y. et al., "Optically Accessible Microchannels Formed in a Single-Crystal Silicon Substrate for Studies of Blood Rheology, " Microvascular Research, vol. 44, 226-240, (1992).

Kobayashi, M. et al., "Fluorescence-based DNA Minisequence Analysis for Detection of Known Single-base Changes in Genomic DNA, " Molecular and Cellular Probes, vol. 9, 175-182 (1995).

Kohsaka, H. and Carson, D., "Solid-Phase Polymerase Chain Reaction, " Journal of Clinical Laboratory Analysis, vol. 8, 452-455 (1994).

Kricka, L. et al., "Imaging of the Chemiluminescent Reactions in Mesoscale Silicon-Glass Microstructures, " J Biolumin Chemilumin, vol. 9, 135-138 (1994).

Kuppuswamy, M. et al., "Angle Nucleotide Primer Extension to Detect Genetic Diseases: Experimental Application to Hemophilia B (Factor IX) and Cystic Fibrosis Genes, " Proc. Natl. Acad. Sci. USA, vol. 88, 1143-1147 (1991).

Lagerkvist, A. et al., "Manifold Sequencing: Efficient Processing of Large Sets of Sequencing Reactions, " Proc. Natl. Acad. Sci. USA, vol. 91, 2245-2249 (1994).

Lamture, J. et al., "Direct Detection of Nucleic Acid Hybridization on the Surface of a Charge Coupled Device, " Nucleic Acids Research, vol. 22, No. 11, 2121-2125 (1994).

Mauro, J. et al., "Fiber-Optic Fluorometric Sensing of Polymerase Chain Reaction-Amplified DNA Using an Immobilized DNA Capture Protein, " Analytical Biochemistry, vol. 235, 61-72 (1996).

Maxam, A. and Gilbert, W., "A New Method for Sequencing DNA, " Proc. Natl. Acad. Sci. USA, vol. 74, No. 2, 560-564 (1977).

Metzker, M. et al., "Termination of DNA Synthesis by Novel 3'-Modified-Deoxyribonucleoside 5'-Triphosphates, " Nucleic Acids Res., vol. 22, No. 20, 4259-4267 (1994).

Nikiforov, T. et al., "Genetic Bit Analysis: A Solid Phase Method for Typing Single Nucleotide Polymorphisms, " Nucleic Acids Res., vol. 22, No. 20, 4167-4175 (1994).

Riccelli, P. and Benight, A., "Tetramethylammonium Does not Universally Neutralize Sequence Dependent DNA Stability, " Nucleic Acids Res., vol. 21, No. 16, 3786-3788 (1993).

Rosenthal, A. et al., "Large-Scale Production of DNA Sequencing Templates by Microtitre Format PCR, " Nucleic Acids Research, vol. 21, No. 1, 173-174 (1993).

Sanger, F. et al., "DNA Sequencing with Chain-Terminating Inhibitors, " Proc. Natl. Acad. Sci. USA, vol. 74, No. 12, 5463-5467 (1977).

Shoffner, M. et al., "Chip PCR. I. Surface Passivation of Microfabricated Silicon-Glass Chips for PCR, " Nucleic Acids Research, vol. 24, No. 2, 375-379 (1996).

Sokolov, B., "Primer Extension Technique for the Detection of Single Nucleotide in Genomic DNA, " Nucleic Acids Res., vol. 18, No. 12, 3671 (1989).

Strezoska, Z. et al., "DNA Sequencing by Hybridization: 100 Bases Read by a Non-Gel-Based Method, " Proc. Natl. Acad. Sci. USA, vol. 88, 10089-10093 (1991).

Syvanen, A. et al., "Convenient and Quantitative Determination of the Frequency of a Mutant Allele Using Solid-Phase Minisequencing: Application to Aspartylglucosaminuria in Finland, " Genomics, vol. 12, 590-595 (1992).

Versalovic, J. et al., "Distribution of Repetitive DNA Sequences in Eubacteria and Application to Fingerprinting of Bacterial Genomes, " Nucleic Acid Res., vol. 19, No. 24, 6823-6831 (1991).

Warren S., "The Expanding World of Trinucleotide Repeats, " Science, vol. 271, 1374-1375 (1996).

Wilding, P. et al., "Manipulation and Flow of Biological Fluids in Straight Channels Micromachined in Silicon, " Clin. Chem., vol. 40, No. 1, 43-47 (1994).

Williams, J., et al., "Studies of Oligonucleotide Interactions by Hybridization to Arrays: The Influence of Dangling Ends on Duplex Yield, " Nucleic Acids Res., vol. 22, No. 8, 1365-1367 (1994).

Woolley, A. and Mathies, R., "Ultra-High-Speed DNA Fragment Separations Using Microfabricated Capillary Array Electrophoresis Chips, " Proc. Natl. Acad. Sci. USA, vol. 91, 11348-11352 (1994).

Lipshutz et al. "Using Oligonucleotide Probe Arrays to Access Genetic Diversity" Biotechniques, vol. 19, No. 3, pp. 442-447, Mar. 1995.

Southern et al. "Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation Using Experimental Models" Genomics, vol. 13, pp. 1008-1017, 1992.

Brenner et al., "DNA fingerprinting by sampled sequencing, " PNAS, 86:8902-8906 (1989).

Fodor et al., "Light-Directed, Spatially Addressable Parallel Chemical Synthesis, " Science, 251:767-773 (1991).

Hoheisel, "Application of hybridization techniques to genome mapping and sequencing, " Trends in Genetics, 10(3):79-83 (1994).

Pease et al., "Light-generated oligonucleotide arrays for rapid DNA sequence analysis, " PNAS, 91:5022-5026 (1994).

Sapolsky et al., "Mapping genomic library clones using oligonucleotide arrays, " Hilton Head Conf., SC, Sept. 17-21, 1994.

Smith, "Ligation-mediated PCR of restriction fragments from large DNA molecules, " PCR Methods and Applications, Cold Spring Harbor Laboratory Press, 2:21-27 (1992).

Szybalski, "Universal resriction endonucleases: designing novel cleavage specificities by combining adapter oligo-deoxynucleotide and enzyme moieties, " Gene, 40:169-173 (1985).

Szybalski et al., "Class-IIs restriction enzymes: a review, " Gene, 100:13-26 (1991).

Unrau, et al., "Non-cloning amplification of specific DNA from whole genomic DNA digests using DNA indexers, " Gene, 145:163-169 (1994).

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, modifications, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features herein before set forth and as follows in the scope of the appended claims. All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

WHAT IS CLAIMED IS:

1. A collection comprising subpopulations of particles, the particles in each subpopulation having one or more characteristics that distinguish the particles of one subpopulation from those of another subpopulation, in which said collection is further characterized as having about 10^3 or more distinct subpopulations of particles.

2. The collection of claim 1 which is further characterized as having about 10^4 or more distinct subpopulations of particles.

3. The collection of claim 1 which is further characterized as having about 10^5 or more distinct subpopulations of particles.

4. The collection of claim 1 which is further characterized as having about 10^6 or more distinct subpopulations of particles.

5. The collection of claim 1 in which the particles in each subpopulation further comprise bound nucleic acid.

6. The collection of claim 5 in which said bound nucleic acid comprises a predetermined polynucleotide sequence.

7. The collection of claim 5 in which said bound nucleic acid comprises an ascertainable polynucleotide sequence.

8. A fluid array comprising:

(a) a collection of subpopulations of particles, the particles in each subpopulation having (i) one or more characteristics that distinguish the particles of one subpopulation from those of another subpopulation and (ii) bound nucleic acid, said collection further characterized as having about 10^3 or more distinct subpopulations of particles, and

(b) a fluid carrier.

9. The fluid array of claim 8 in which said fluid carrier comprises a liquid.

10. The fluid array of claim 8 in which said fluid carrier comprises a gas.

11. The fluid array of claim 8 in which the subpopulations of particles of said collection are substantially comingled with one another.

12. The fluid array of claim 8 in which the subpopulations of particles of said collection are substantially segregated from one another.

13. The fluid array of claim 8 in which said one or more characteristics include a distinctive fluorescence emission signature.

14. The fluid array of claim 8 in which said nucleic acid bound to the particles of one subpopulation differs from that bound to the particles of another subpopulation.

15. The fluid array of claim 8 in which said bound nucleic acid comprises a predetermined polynucleotide sequence.

16. The fluid array of claim 8 in which said bound nucleic acid comprises an ascertainable polynucleotide sequence.

17. The fluid array of claim 8 in which said collection is further characterized as having about 10^4 or more distinct subpopulations of particles.

18. The fluid array of claim 8 in which said collection is further characterized as having about 10^5 or more distinct subpopulations of particles.

19. The fluid array of claim 8 in which said collection is further characterized as having about 10^6 or more distinct subpopulations of particles.

20. The fluid array of claim 8 in which said bound nucleic acid comprises DNA.

21. The fluid array of claim 8 in which said bound nucleic acid comprises RNA.

22. The fluid array of claim 1 in which said bound nucleic acid comprises 9 or more nucleotide residues.

23. The fluid array of claim 1 in which said bound nucleic acid comprises 18 or more nucleotide residues.

24. The fluid array of claim 1 in which said bound nucleic acid comprises 36 or more nucleotide residues.

25. A composition of matter comprising a solid particle including (i) bound nucleic acid having a known polynucleotide sequence, (ii) a label comprising a dye that exhibits a distinctive fluorescent emission signature, and (iii) a substance that, in the absence of an analyte of interest comprising a polynucleotide sequence substantially complementary to said known polynucleotide sequence, can quench the fluorescence emission of said dye.

26. A method of characterizing a nucleic acid of interest comprising the steps of:

(a) providing a plurality of oligomer probes of known or ascertainable sequence, bound to a respective number of subpopulations of particles having one or more charactetistics that distinguish the particles of one subpopulation from those of another subpopulation so that the sequence of a probe is identifiable according to the unique characteristic of the particular subpopulation of particles;

(b) hybridizing said oligomer probes with the nucleic acid of interest to obtain complementary complexes, and;

(c) determining the sequence of the nucleic acid of interest in said complementary complexes by referring to the unique characteristic associated with each subpopulation of particles carrying the probe of known or ascertainable sequence.

27. The method according to claim 26 wherein said oligomer probes further comprise a fluorescent reporter molecule, whereby the fluorescence signal from said reporter molecule changes as a function of hybridization complementarity between the probe and the nucleic acid sequence of interest.

28. The method according to claim 26 wherein said oligomer probes of each said set comprise contiguous overlapping probes differing from each other by at least one base selected from A, C, T, or G.

29. The method according to claim 28 wherein contiguous overlapping probes contain degenerate sequences.

30. The method of claim 26 wherein said oligomer probes are selected from the group consisting of DNA, RNA, PNA, and modifications thereof.

31. The method according to claim 26 wherein the nucleic acid of interest comprises at least one mutation or a set of mutations linked to a clinical condition or a predisposition to said clinical condition, wherein said clinical condition or predisposition thereto is selected from a group consisting of hereditary diseases, neural diseases, muscle and bone diseases, malignant diseases, infectious diseases, metabolic diseases, and combination thereof.

32. The method according to claim 26 whereby said method is useful for determining a genetic distance between the nucleic acid of interest and a reference sample.

33. The method according to claim 26 wherein the nucleic acid of interest is amplified by nucleic acid amplification methods.

34. The method according to claim 26 whereby the nucleic acid of interest is analyzed by a primer extension reaction.

35. A method of quantitating an analyte of interest in a sample comprising the steps of:

(a) contacting said sample with a detectable probe bound to a fluorescently addressable particle; and

(b) measuring the quantity of said analyte by comparing to a standard curve, whereby the standard curve comprises values from at least two known quantities of a reference analyte.

36. An array of nucleic acid probes wherein each of said probes is bound to a discrete fluorescently addressable set of microparticles, wherein each said set is positioned in a predetermined well of a microtiter plate.

37. The array of claim 36 whereby said array has between 65 and 1,000,000 oligonucleotide probes.

38. A library of oligonucleotide probes of known sequence in which each discrete probe is bound to a respective fluorescent microparticle stained with two or more fluorescent dyes and each said dye has the potential of having at least eight different levels of fluorescence intensity.

39. The library of claim 38 wherein each said probe comprises a sequence composed of randomly assembled nucleotides, whereby the total number of probes corresponds to formula 4^N, N being the number of nucleotides in the probe.

40. The library of claim 38 wherein said sequence of each said probe has the potential to correspond to any of naturally occurring amino acids.

41. A device for identifying an analyte of interest among a plurality of different analytes in a sample, said device comprising a fluorescently addressable microparticle having on its surface at least one bound probe of known sequence labeled with a fluorescent reporter dye, to which said analyte of interest binds in complementary fashion so that said fluorescent reporter dye on the binding probe undergoes a change in fluorescence output indicating the presence of the analyte in the sample and said analyte is identified according to the fluorescent signature of the microparticle.

42. A method for constructing a library of oligomer probes of known sequence comprising the steps of:

(a) coupling each of four bases, A, C, G, and T, to at least four respective sets of fluorescently distinguishable microparticles;

(b) stacking by means of nucleotide synthesis chemistry to the microparticle-coupled base the next base selected from A, C, G, or T;

(c) sorting microparticles according to the formed sequence, and;

(d) repeating the nucleotide synthesis and sorting steps (b) and (c) until the desired sequence of the oligomer probe is obtained.

43. The method according to claim 42 wherein said sorting is performed by flow cytometry.

44. A method for constructing a library of oligomer probes of known sequence comprising the steps of:

(a) synthesizing by means of nucleotide synthesis chemistry N number of sets of oligomer probes of desired sequence, and;

(b) coupling at least one oligomer probe from one of said N number of sets of oligomer probes to a respective set of fluorescently distinct microparticles labeled with at least two fluorescent dyes having at least eight different levels of fluorescence intensity.

45. An array of nucleic acid probes comprising a plurality of fluorescently addressable microparticles, each stained with at least two fluorescent dyes and carrying at least one distinct nucleic acid probe, microparticles being arrayed in a two-dimensional pattern over a plane of a microtiter plate.

46. The array of claim 45 wherein said microtiter plate has between about 96 and 2,034 reaction wells.

47. The array of claim 45 whereby said array has more than 64 nucleic acid probes.

48. The array of claim 45 wherein said nucleic acid probes comprise degenerate sequences wherein R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, and N=A/C/G/T.

49. The method for carrying out a sequencing by hybridization, an analysis of a gene expression by hybridization of gene-specific mRNA or cDNA to an array of complementary probes, and quantitation of copies of nucleic acid sequences of interest by comparing to a known quantity of a reference material, said method involving the use of the array of claim 45.

50. The array of claim 45, which is useful for screening molecules that bind to nucleic acids of said array wherein said molecules have various types of biological activities comprising hormonal, neurotransmitter, metabolic, genetic, pharmacologic, immunologic, pathologic, toxic, and anti-mitotic activities.

51. A liquid array, comprising: a mixture of sets of fluorescently addressable microspheres in a liquid,

wherein each set has a distinct fluorescent signature and wherein each set is conjugated with a different oligonucleotide probe, whereby detection of a fluorescent signature identifies the oligonucleotide probe; said array having at least 100 probes and no more than 1,000,000 probes of about 9 to 20 nucleotides in length; said array comprising at least four groups of microspheres, wherein a first group is exactly complementary to a reference sequence and comprises probes that completely span the reference sequence and, relative to the reference sequence, overlap one another in sequence, and wherein three additional groups of microspheres, each of which is identical to said first group but at least one different nucleotide, which different nucleotide is located in the same position in each of the three additional sets but which is a different nucleotide in each set.

52. An enzymatic process for analyzing a nucleic acid sequence present in a sample of interest, comprising the steps of:

a) providing an array of fluorescently addressable microparticles stained with at least two distinct fluorescent dyes;

b) hybridizing the nucleic acid in the sample of interest with said array; and,

c) analyzing the obtained hybrid by a primer extension enzymatic process.

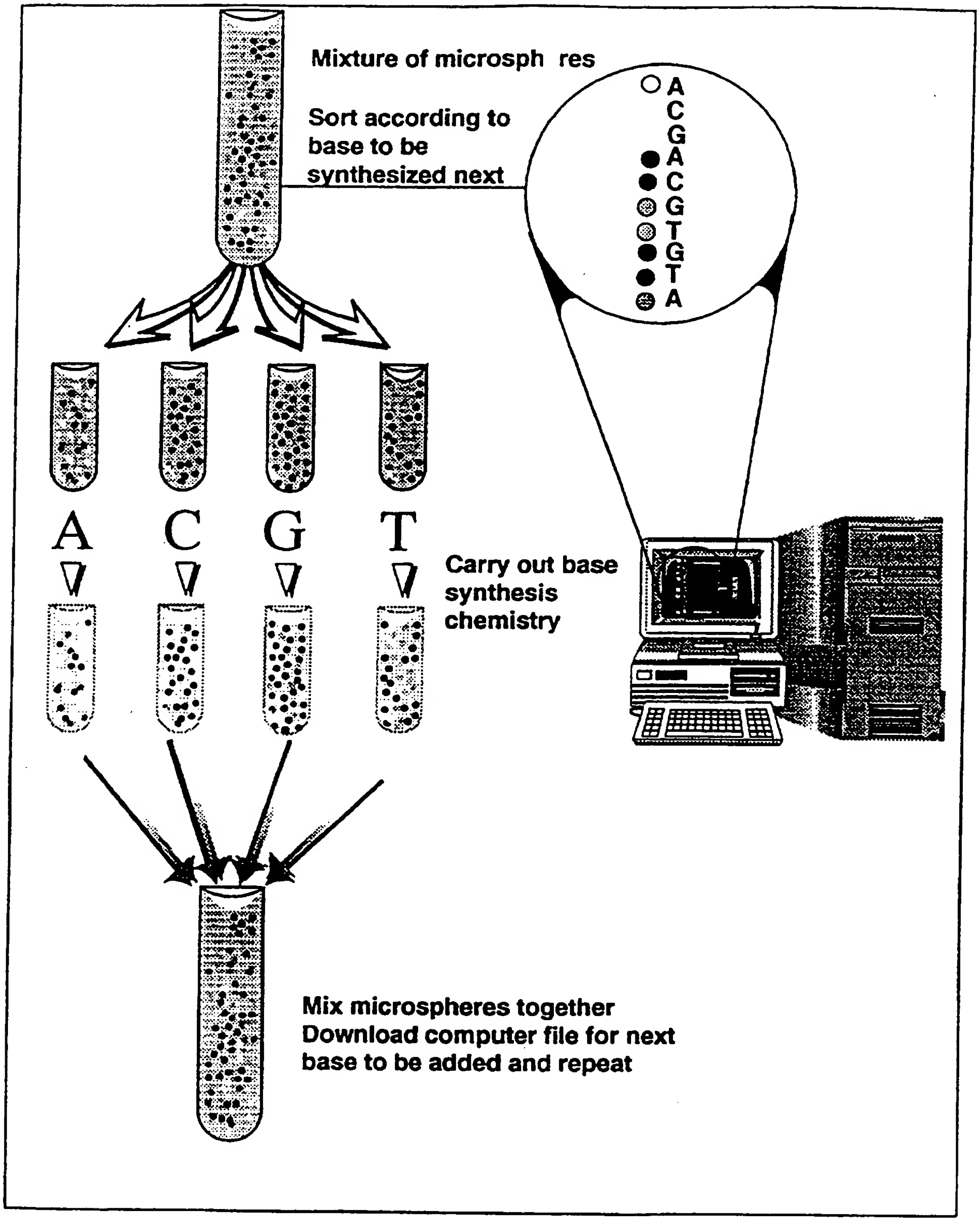
Mixture of microsph res
Sort according to base to be synthesized next
A
C
G
A
C
G
T
G
T
A
A
C
G
T
Carry out base synthesis chemistry
Mix microspheres together
Download computer file for next base to be added and repeat

Figure 1. Segmental synthesis of oligonucleotides by sorting.